

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 30, 2004, 10:50:47 ; Search time 4.87938 Seconds  
(without alignments)  
1073.493 Million cell updates/sec

Title: US-09-989-481-4  
Perfect score: 192  
Sequence: 1 LGTFWGDTLNCWMLSAFSRYARCLAEGHDGPTQ 33

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A\_Geneseq\_19Jun03:\*  
1: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1980.DAT:\*  
2: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1981.DAT:\*  
3: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1982.DAT:\*  
4: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1983.DAT:\*  
5: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1984.DAT:\*  
6: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1985.DAT:\*  
7: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1986.DAT:\*  
8: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1987.DAT:\*  
9: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1988.DAT:\*  
10: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1989.DAT:\*  
11: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1990.DAT:\*  
12: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1991.DAT:\*  
13: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1992.DAT:\*  
14: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1993.DAT:\*  
15: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1994.DAT:\*  
16: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1995.DAT:\*  
17: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1996.DAT:\*  
18: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1997.DAT:\*  
19: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1998.DAT:\*  
20: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:\*  
21: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:\*  
22: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:\*  
23: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2002.DAT:\*  
24: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2003.DAT:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,

and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	%		Query			DB	ID	Description
	Score	Match	Length	Match	Length			
1	192	100.0	33	19	AAW59046			Human MNTF1-F6 pro
2	56.5	29.4	584	21	AAW90949			Comamonas testoste
3	55.5	28.9	227	22	AAU19525			Human diagnostic a
4	55	28.6	694	23	ABB93813			Herbicidally activ
5	53.5	27.9	326	21	AAB23939			Hepatitis B virus
6	53.5	27.9	384	22	AAB94802			Human protein sequ
7	53	27.6	794	22	ABG25667			Novel human diagno
8	52	27.1	563	12	AAR10330			Gene product with
9	51	26.6	56	22	ABG60253			Human ovarian anti
10	51	26.6	56	22	AAM94423			Human reproductive
11	51	26.6	56	23	ABG61724			Novel ovarian rela
12	51	26.6	466	23	AAE23630			Escherichia coli 6
13	51	26.6	474	22	AAU34677			E. coli cellular p
14	51	26.6	1207	22	AAM78524			Human protein SEQ
15	51	26.6	1207	22	AAB84604			Amino acid sequenc
16	51	26.6	1207	23	ABP54791			Human epidermal gr
17	51	26.6	1222	22	ABB11946			Human precursor pr
18	51	26.6	1222	22	AAM79508			Human protein SEQ
19	51	26.6	1225	22	ABG24444			Novel human diagno
20	51	26.6	1258	22	ABG24819			Novel human diagno
21	50	26.0	726	22	AAB68590			AtCNGC2/DND1 prote
22	50	26.0	726	23	ABB93489			Herbicidally activ
23	50	26.0	844	22	ABB61902			Drosophila melanog
24	49.5	25.8	842	23	ABB06211			HIV Env isolate SF
25	49.5	25.8	842	24	ABU66565			Human immunodefici
26	49.5	25.8	847	21	AAU97073			Variant HIV-1 SF16
27	49.5	25.8	1054	22	ABB60410			Drosophila melanog
28	49	25.5	414	23	ABB54212			Lactococcus lactis
29	49	25.5	665	19	AAW54425			Human PS112 protei
30	49	25.5	665	21	AAB08415			Protein encoded by
31	48.5	25.3	150	22	AAU87494			Novel central nerv
32	48	25.0	46	24	ABP57309			Kazal type protein
33	48	25.0	94	22	ABG19025			Novel human diagno
34	48	25.0	113	17	AAR96565			Hepatitis C virus
35	48	25.0	116	22	ABG30144			Novel human diagno
36	48	25.0	526	22	AAG78608			Lawsonia intracell
37	48	25.0	731	22	ABG30155			Novel human diagno
38	48	25.0	743	23	ABB92349			Herbicidally activ
39	48	25.0	839	23	AAE23384			Human intracellula
40	47.5	24.7	84	22	ABG53718			Human liver peptid
41	47.5	24.7	84	22	ABB38826			Peptide #6332 enco
42	47.5	24.7	84	22	ABB23847			Protein #5846 enco
43	47.5	24.7	84	22	AAM59470			Human brain expres
44	47.5	24.7	84	22	AAM72034			Human bone marrow
45	47.5	24.7	84	22	AAM19413			Peptide #5847 enco

# ALIGNMENTS

RESULT 1

AAW59046

ID AAW59046 standard; Protein; 33 AA.

XX

AC AAW59046;

XX

DT 11-AUG-1998 (first entry)

XX

DE Human MNTF1-F6 protein fragment.

XX

KW Motoneuronotrophic factor; MNTF-1; MNTF1-F6; human; axon regeneration;  
KW motoneuron; diagnose; treatment; disease; wound healing; scar tissue;  
KW keloid.

XX

OS Homo sapiens.

XX

PN WO9813492-A2.

XX

PD 02-APR-1998.

XX

PF 22-SEP-1997; 97WO-US17142.

XX

PR 12-SEP-1997; 97US-0928862.

PR 27-SEP-1996; 96US-0026792.

PR 15-NOV-1996; 96US-0751225.

XX

PA (KMBI-) KM BIOTECH INC.

XX

PI Chau RMW;

XX

DR WPI; 1998-230703/20.

DR N-PSDB; AAV11748.

XX

PT Motoneurotrophic factor MNTF1-F3 and MNTF1-F6 - useful for  
PT motoneuron regeneration, diagnosing or treating motoneuron disease  
PT and to accelerate wound healing without scar formation

XX

PS Claim 4; Fig 2B; 78pp; English.

XX

CC This sequence, represents a fragment of a novel human motoneurotrophic  
CC factor, MNTF1-F6. Such factors are used to promote regeneration of the  
CC axon of a motoneurone, to diagnose and treat motoneurone disease in a  
CC mammal or to accelerate wound healing whilst concomitantly minimising  
CC or inhibiting scar tissue and/or keloid formation in an area associated  
CC with a wound. For promoting axonal regeneration, the polypeptide is  
CC administered at a concentration of 5 ng-50 mg, whereas for inhibiting  
CC hereditary motoneurone disease, the dosage is 5-100 (especially 30-50)ng  
CC per kg body weight.

XX

SQ Sequence 33 AA;

Query Match 100.0%; Score 192; DB 19; Length 33;

Best Local Similarity 100.0%; Pred. No. 2.3e-18;

Matches 33; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LGTFWGDTLNCWMLSAFSRYARCLAEGHDGPTQ 33

|||||

Db 1 LGTFWGDTLNCWMLSAFSRYARCLAEGHDGPTQ 33

RESULT 2

AAW90949

ID AAW90949 standard; Protein; 584 AA.

XX

AC AAW90949;

XX

DT 21-JUL-2000 (first entry)

XX

DE Comamonas testosteroni R5 phenol hydroxylase protein #7.

XX

KW Phenol hydroxylase; microbe; phenol decomposition.

XX

OS Comamonas testosteroni.

XX

PN JP2000069968-A.

XX

PD 07-MAR-2000.

XX

PF 28-AUG-1998; 98JP-0243249.

XX

PR 28-AUG-1998; 98JP-0243249.

XX

PA (KAIY-) KAIYO BIOTECHNOLOGY KENKYUSHO KK.

XX

DR WPI; 2000-264445/23.

DR N-PSDB; AAA11713.

XX

PT Structural gene and a regulator gene of phenol hydroxylase of Comamonas  
PT testosteroni R5 - used for decomposing phenol

XX

PS Claim 2a; Page 15-16; 18pp; Japanese.

XX

CC This invention describes a novel microbe for decomposing phenol which  
CC carries a phenol hydroxylase protein. This sequence represents a  
CC phenol hydroxylase protein encoded by the sequence represented in  
CC AAA11713 which is described in the method of the invention.

XX

SQ Sequence 584 AA;

Query Match 29.4%; Score 56.5; DB 21; Length 584;  
Best Local Similarity 33.3%; Pred. No. 42;  
Matches 18; Conservative 3; Mismatches 6; Indels 27; Gaps 4;

QY 5 WG--DTLNCWML-----SAFSR-----YARCLAEG---HDGP 31  
|| | :||| ||| | :| || | | : |  
Db 155 WGPQDQPSCWMLLG YASGYSSAFFRRPVFFKEMQCSTCGHAHCLIEGRFQHEWP 208

RESULT 3

AAU19525

ID AAU19525 standard; Protein; 227 AA.

XX

AC AAU19525;

XX

DT 04-DEC-2001 (first entry)  
 XX  
 DE Human diagnostic and therapeutic polypeptide (DITHP) #111.  
 XX  
 KW Human; receptor; diagnostic; therapeutic; gene therapy; vaccine;  
 KW cell proliferative disorder; Crohn's disease; lymphoma; leukaemia;  
 KW acquired immune deficiency syndrome; AIDS; autoimmune disorder;  
 KW respiratory disorder.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200162927-A2.  
 XX  
 PD 30-AUG-2001.  
 XX  
 PF 21-FEB-2001; 2001WO-US06059.  
 XX  
 PR 24-FEB-2000; 2000US-0184693.  
 PR 24-FEB-2000; 2000US-0184697.  
 PR 24-FEB-2000; 2000US-0184698.  
 PR 24-FEB-2000; 2000US-0184768.  
 PR 24-FEB-2000; 2000US-0184769.  
 PR 24-FEB-2000; 2000US-0184770.  
 PR 24-FEB-2000; 2000US-0184771.  
 PR 24-FEB-2000; 2000US-0184772.  
 PR 24-FEB-2000; 2000US-0184773.  
 PR 24-FEB-2000; 2000US-0184774.  
 PR 24-FEB-2000; 2000US-0184776.  
 PR 24-FEB-2000; 2000US-0184777.  
 PR 24-FEB-2000; 2000US-0184797.  
 PR 24-FEB-2000; 2000US-0184813.  
 PR 24-FEB-2000; 2000US-0184837.  
 PR 24-FEB-2000; 2000US-0184841.  
 PR 24-FEB-2000; 2000US-0185213.  
 PR 24-FEB-2000; 2000US-0185216.  
 PR 12-MAY-2000; 2000US-0203785.  
 PR 15-MAY-2000; 2000US-0204226.  
 PR 16-MAY-2000; 2000US-0204525.  
 PR 16-MAY-2000; 2000US-0204821.  
 PR 16-MAY-2000; 2000US-0204908.  
 PR 16-MAY-2000; 2000US-0205232.  
 PR 17-MAY-2000; 2000US-0204815.  
 PR 17-MAY-2000; 2000US-0204863.  
 PR 17-MAY-2000; 2000US-0205221.  
 PR 17-MAY-2000; 2000US-0205285.  
 PR 17-MAY-2000; 2000US-0205286.  
 PR 17-MAY-2000; 2000US-0205287.  
 PR 17-MAY-2000; 2000US-0205323.  
 PR 17-MAY-2000; 2000US-0205324.  
 XX  
 PA (INCY-) INCYTE GENOMICS INC.  
 XX  
 PI Panzer SR, Spiro PA, Banville SC, Shah P, Chalup MS, Chang SC;  
 PI Chen A, D'Sa SA, Amshey S, Dahl CR, Dam TC, Daniels SE;  
 PI Dufour GE, Flores V, Fong WT, Greenawalt LB, Hillman JL, Jones AL;  
 PI Liu TF, Roseberry AM, Rosen BH, Russo FD, Stockdreher TK, Daffo A;  
 PI Wright RJ, Yap PE, Yu JY, Bradley DL, Bratcher SR, Chen W;

PI Cohen HJ, Hodgson DM, Lincoln SE, Jackson S;  
XX  
DR WPI; 2001-502867/55.  
DR N-PSDB; AAS31096.  
XX  
PT Polynucleotides encoding diagnostic and therapeutic proteins, e.g.  
PT enzymes, hormones and receptors, useful in diagnostics and therapeutics  
PT -

XX  
PS Claim 27; Page 464; 522pp; English.  
XX

CC The invention relates to polynucleotides (I) encoding diagnostic and  
CC therapeutic (DITHP) polypeptides (II), which include e.g. enzymes,  
CC and proteins involved in growth and development and receptors. (I) and  
CC (II) may be used in the prevention, diagnosis and treatment of diseases  
CC associated with inappropriate DITHP expression. For example, (I) and  
CC (II) may be used to treat disorders associated with decreased polypeptide  
CC expression by rectifying mutations or deletions in a patient's genome,  
CC that affect the activity of the DITHPs, by expressing inactive proteins  
CC or supplementing the patient's own production of them. (I) and (II)  
CC may be used to treat diseases, for example, cell proliferative disorder,  
CC Crohn's disease, acquired immune deficiency syndrome (AIDS), lymphoma,  
CC leukaemia, autoimmune disorders, and respiratory disorders. Additionally,  
CC (I) may be used to produce the DITHPs, by inserting the nucleic acids  
CC into a host cell and culturing the cell to express the protein. (I) and  
CC its complementary sequences may also be used as DNA probes in diagnostic  
CC assays to detect and quantitate the presence of similar nucleic acids in  
CC samples, and therefore which patients may be in need of restorative  
CC therapy. (II) may also be used as antigens in the production of  
CC antibodies against DITHPs and in assays to identify modulators of DITHP  
CC expression and activity. The anti-DITHP antibodies and antagonists may  
CC also be used to down regulate expression and activity. The anti-DITHP  
CC antibodies may also be used as diagnostic agents for detecting the  
CC presence of DITHPs in samples (e.g. by enzyme linked immunosorbant  
CC assay (ELISA)). AAU19415-AAU19625 represent human diagnostic and  
CC therapeutic (DITHP) polypeptides of the invention.

XX  
SQ Sequence 227 AA;

Query Match 28.9%; Score 55.5; DB 22; Length 227;  
Best Local Similarity 22.6%; Pred. No. 21;  
Matches 14; Conservative 5; Mismatches 8; Indels 35; Gaps 2;

Qy 4 FWGDTLNCW-----MLSAFSRY-----ARCLAEGH 28  
||| ||| : |||| : : |:: |  
Db 125 FWGGQRNCWGSRSRASAPLFSAFSEFFPAFGGVFSSFDTGFRSFGSLGSGGLSSFCMSYGS 184

Qy 29 DG 30  
||  
Db 185 DG 186

RESULT 4  
ABB93813  
ID ABB93813 standard; Protein; 694 AA.  
XX  
AC ABB93813;

XX  
 DT 31-MAY-2002 (first entry)  
 XX  
 DE Herbicidally active polypeptide SEQ ID NO 3024.  
 XX  
 KW Herbicidal; plant; agriculture; herbicide.  
 XX  
 OS Arabidopsis thaliana.  
 XX  
 PN WO200210210-A2.  
 XX  
 PD 07-FEB-2002.  
 XX  
 PF 28-AUG-2001; 2001WO-EP09892.  
 XX  
 PR 28-AUG-2001; 2001WO-EP09892.  
 XX  
 PA (FARB ) BAYER AG.  
 XX  
 PI Tietjen K, Weidler M;  
 XX  
 DR WPI; 2002-269010/31.  
 XX  
 PT Identifying plant target proteins for herbicidally active compounds,  
 PT comprising aligning and comparing nucleic acid or amino acid sequences  
 PT from plant with nucleic acid or amino acid sequences from non-plant  
 PT organisms -  
 XX  
 PS Claim 5; SEQ ID NO 3024; 261pp + Sequence Listing; English.  
 XX  
 CC The invention relates to identifying target proteins  
 CC (ABB90790-ABB94016) for herbicidally active compounds, comprising  
 CC aligning and comparing nucleic acid or amino acid sequences from plant  
 CC with nucleic acid or amino acid sequences from non-plant organisms using  
 CC suitable search parameters, where plant sequences having an E-value  
 CC greater by a factor of 3 than the E-value of most similar non-plant  
 CC sequences are selected. The polypeptides or nucleic acids encoding them  
 CC are useful for identifying modulators. The identified modulators are  
 CC useful as herbicides.  
 XX  
 SQ Sequence 694 AA;  
  
 Query Match 28.6%; Score 55; DB 23; Length 694;  
 Best Local Similarity 35.9%; Pred. No. 80;  
 Matches 14; Conservative 2; Mismatches 9; Indels 14; Gaps 2;  
  
 QY 2 GT-FWGDTLN-----CWMLSAFSRYARCLAE 26  
 || :|| || || | | :|| |  
 Db 255 GTVWWGIALNMIAYFVAHAAGACWYLLGVQRSAKCLKE 293

RESULT 5  
 AAB23939  
 ID AAB23939 standard; Protein; 326 AA.  
 XX  
 AC AAB23939;  
 XX

DT 18-JAN-2001 (first entry)  
 XX  
 DE Hepatitis B virus protein bound arrestin protein sequence SEQ ID NO:2.  
 XX  
 KW Hepatitis B virus; HBV; arrestin; binding.  
 XX  
 OS Hepatitis B virus.  
 XX  
 PN CN1257919-A.  
 XX  
 PD 28-JUN-2000.  
 XX  
 PF 21-DEC-1998; 98CN-0125693.  
 XX  
 PR 21-DEC-1998; 98CN-0125693.  
 XX  
 PA (UYFU-) UNIV FUDAN.  
 XX  
 PI Yu L, Wang X, Fu Q;  
 XX  
 DR WPI; 2000-544292/50.  
 DR N-PSDB; AAA99087.  
 XX  
 PT Hepatitis B virus protein bound arrestin -  
 XX  
 PS Claim 1; Page 13; 16pp; Chinese.  
 XX  
 CC The present sequence represents a specifically claimed protein  
 CC sequence from the present invention. The present invention describes  
 CC Hepatitis B virus (HBV) protein bound arrestin. Also described is a  
 CC method for the preparation of the novel protein and polynucleotide of  
 CC the invention.  
 XX  
 SQ Sequence 326 AA;  
  
 Query Match 27.9%; Score 53.5; DB 21; Length 326;  
 Best Local Similarity 40.0%; Pred. No. 56;  
 Matches 12; Conservative 6; Mismatches 11; Indels 1; Gaps 1;  
  
 Qy 4 FWGDTLNCWMLSAFSRYARCLAEHGDGPTQ 33  
 || ||:: : || | :|:| | ||:  
 Db 87 FWCDTIHTGVYPILSRSLRQMAQGD-PTE 115

RESULT 6

AAB94802

ID AAB94802 standard; Protein; 384 AA.

XX

AC AAB94802;

XX

DT 26-JUN-2001 (first entry)

XX

DE Human protein sequence SEQ ID NO:15935.

XX

KW Human; primer; detection; diagnosis; antisense therapy; gene therapy.

XX

OS Homo sapiens.



XX  
 PN EP1074617-A2.  
 XX  
 PD 07-FEB-2001.  
 XX  
 PF 28-JUL-2000; 2000EP-0116126.  
 XX  
 PR 29-JUL-1999; 99JP-0248036.  
 PR 27-AUG-1999; 99JP-0300253.  
 PR 11-JAN-2000; 2000JP-0118776.  
 PR 02-MAY-2000; 2000JP-0183767.  
 PR 09-JUN-2000; 2000JP-0241899.  
 XX  
 PA (HELI-) HELIX RES INST.  
 XX  
 PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;  
 PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;  
 XX  
 DR WPI; 2001-318749/34.  
 XX  
 PT Primer sets for synthesizing polynucleotides, particularly the 5602  
 PT full-length cDNAs defined in the specification, and for the detection  
 PT and/or diagnosis of the abnormality of the proteins encoded by the  
 PT full-length cDNAs -  
 XX  
 PS Claim 8; SEQ ID 15935; 2537pp + CD ROM; English.  
 XX  
 CC The present invention describes primer sets for synthesising 5602  
 CC full-length cDNAs defined in the specification. Where a primer set  
 CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary  
 CC to the complementary strand of a polynucleotide which comprises one of  
 CC the 5602 nucleotide sequences defined in the specification, where the  
 CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination  
 CC of an oligonucleotide comprising a sequence complementary to the  
 CC complementary strand of a polynucleotide which comprises a 5'-end  
 CC sequence and an oligonucleotide comprising a sequence complementary to a  
 CC polynucleotide which comprises a 3'-end sequence, where the  
 CC oligonucleotide comprises at least 15 nucleotides and the combination of  
 CC the 5'-end sequence/3'-end sequence is selected from those defined in  
 CC the specification. The primer sets can be used in antisense therapy and  
 CC in gene therapy. The primers are useful for synthesising polynucleotides,  
 CC particularly full-length cDNAs. The primers are also useful for the  
 CC detection and/or diagnosis of the abnormality of the proteins encoded by  
 CC the full-length cDNAs. The primers allow obtaining of the full-length  
 CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and  
 CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to  
 CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632  
 CC represent oligonucleotides, all of which are used in the exemplification  
 CC of the present invention.  
 XX  
 SQ Sequence 384 AA;

Query Match 27.9%; Score 53.5; DB 22; Length 384;  
 Best Local Similarity 40.0%; Pred. No. 67;  
 Matches 12; Conservative 6; Mismatches 11; Indels 1; Gaps 1;

Qy 4 FWGDTLNCWMLSAFSRYARCLAEGHDGPTQ 33

Db                    || ||:: :    || | :|:| | ||:  
87 FWCDTIHTGVYPILSRSLRQMAQKGD-PTE 115

RESULT 7

ABG25667

ID    ABG25667 standard; Protein; 794 AA.

XX

AC    ABG25667;

XX

DT    18-FEB-2002    (first entry)

XX

DE    Novel human diagnostic protein #25658.

XX

KW    Human; chromosome mapping; gene mapping; gene therapy; forensic;

KW    food supplement; medical imaging; diagnostic; genetic disorder.

XX

OS    Homo sapiens.

XX

PN    WO200175067-A2.

XX

PD    11-OCT-2001.

XX

PF    30-MAR-2001; 2001WO-US08631.

XX

PR    31-MAR-2000; 2000US-0540217.

PR    23-AUG-2000; 2000US-0649167.

XX

PA    (HYSE-) HYSEQ INC.

XX

PI    Drmanac RT,    Liu C,    Tang YT;

XX

DR    WPI; 2001-639362/73.

DR    N-PSDB; AAS89854.

XX

PT    New isolated polynucleotide and encoded polypeptides, useful in  
PT    diagnostics, forensics, gene mapping, identification of mutations  
PT    responsible for genetic disorders or other traits and to assess  
PT    biodiversity -

XX

PS    Claim 20; SEQ ID No 56026; 103pp; English.

XX

CC    The invention relates to isolated polynucleotide (I) and  
CC    polypeptide (II) sequences. (I) is useful as hybridisation probes,  
CC    polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
CC    and gene mapping, and in recombinant production of (II). The  
CC    polynucleotides are also used in diagnostics as expressed sequence tags  
CC    for identifying expressed genes. (I) is useful in gene therapy techniques  
CC    to restore normal activity of (II) or to treat disease states involving  
CC    (II). (II) is useful for generating antibodies against it, detecting or  
CC    quantitating a polypeptide in tissue, as molecular weight markers and as  
CC    a food supplement. (II) and its binding partners are useful in medical  
CC    imaging of sites expressing (II). (I) and (II) are useful for treating  
CC    disorders involving aberrant protein expression or biological activity.  
CC    The polypeptide and polynucleotide sequences have applications in  
CC    diagnostics, forensics, gene mapping, identification of mutations  
CC    responsible for genetic disorders or other traits to assess biodiversity

CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. ABG00010-ABG30377 represent novel human  
CC diagnostic amino acid sequences of the invention.  
CC Note: The sequence data for this patent did not appear in the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX

SQ Sequence 794 AA;

Query Match 27.6%; Score 53; DB 22; Length 794;  
Best Local Similarity 40.7%; Pred. No. 1.7e+02;  
Matches 11; Conservative 3; Mismatches 13; Indels 0; Gaps 0;

Qy 7 DTLCNCWMLSAFSRYARCLAEGHDGPTQ 33  
| | | : | : | | | | | | :  
Db 739 DALNLFPLQINPHFTNALAEGHKGETR 765

#### RESULT 8

AAR10330

ID AAR10330 standard; Protein; 563 AA.

XX

AC AAR10330;

XX

DT 25-MAR-2003 (updated)

DT 05-APR-1991 (first entry)

XX

DE Gene product with lipase activity.

XX

KW ATCC 34614.

XX

OS Geotrichum candidum.

XX

PN JP02299588-A.

XX

PD 11-DEC-1990.

XX

PF 27-MAR-1989; 89JP-0074721.

XX

PR 27-MAR-1989; 89JP-0074721.

XX

PA (KURK ) KURITA WATER IND LTD.

PA (OSAO ) OSAKA CITY.

XX

DR WPI; 1991-027567/04.

DR N-PSDB; AAQ10313.

XX

PT Gene for coding protein with lipase activity - is prep'd. from  
PT messenger ribonucleic acid of geo-trichum candidum ATCC 34614

XX

PS Claim 1; Fig 4; 12pp; Japanese.

XX

CC The gene product may be isolated from a transformed expression  
CC sytem, and may be enhanced with stability in heat, alkalai, acid  
CC and organic solvent by position-specific modulation.  
CC (Updated on 25-MAR-2003 to correct PA field.)

XX

SQ Sequence 563 AA;

Query Match 27.1%; Score 52; DB 12; Length 563;  
Best Local Similarity 40.0%; Pred. No. 1.6e+02;  
Matches 14; Conservative 1; Mismatches 14; Indels 6; Gaps 1;

QY 1 LGTFWGDTL-----NCWMLSAFSRYARCLAEGHD 29  
|||| | | | | | ||: || | ||  
Db 478 LGTFHGSDLLFQYYAGPWSSSAYRRYFISFANHHD 512

RESULT 9

ABG60253

ID ABG60253 standard; Protein; 56 AA.

XX

AC ABG60253;

XX

DT 13-AUG-2002 (first entry)

XX

DE Human ovarian antigen #15.

XX

KW Human; ovarian antigen; ovary disorder; breast disorder;  
KW neoplastic disorder; cancer; infectious disease; inflammatory disease;  
KW reproductive system disorder; autoimmune disorder; Alzheimer's disease;  
KW blood-related disorder; hyperproliferative disorder; hair loss;  
KW urinary system disorder; cardiovascular disorder; arrhythmia;  
KW respiratory disorder; musculoskeletal system disorder;  
KW neural activity disorder; neurological disorder; endocrine disorder;  
KW gastrointestinal disorder; liver disorder; pancreatic disorder;  
KW gall bladder disorder; large intestine disorder; developmental disorder;  
KW inherited disorder; wound healing; skin aging; food additive;  
KW preservative.

XX

OS Homo sapiens.

XX

PN WO200155329-A2.

XX

PD 02-AUG-2001.

XX

PF 17-JAN-2001; 2001WO-US01360.

XX

PR 31-JAN-2000; 2000US-0179065.

PR 04-FEB-2000; 2000US-0180628.

PR 07-JUN-2000; 2000US-0209467.

PR 14-SEP-2000; 2000US-0232398.

PR 17-NOV-2000; 2000US-0249300.

PR 01-DEC-2000; 2000US-0250160.

PR 08-DEC-2000; 2000US-0251868.

PR 08-DEC-2000; 2000US-0251990.

XX

PA (HUMA-) HUMAN GENOME SCI INC.

XX

PI Rosen CA, Barash SC, Ruben SM;

XX

DR WPI; 2001-476195/51.

DR N-PSDB; ABK72056.

XX

PT Novel isolated human ovarian related polypeptide useful for  
PT diagnosis/treatment of disorders of ovary and breast such as neoplastic  
PT disorders, infectious diseases, inflammatory diseases, and reproductive  
PT disorders -

XX

PS Claim 11; SEQ ID No 83; 524pp; English.

XX

CC The invention relates to isolated ovarian related polypeptide (ovarian  
CC antigen) comprising a sequence at least 90% identical to a sequence  
CC selected from a polypeptide fragment, domain, epitope or full length  
CC protein of a sequence (S1) appearing as ABG60239-ABG60296 having  
CC biological activity, or a variant, allelic variant or species homologue  
CC of S1. Also included are the cDNA clones encoding the proteins of S1.  
CC S1, an anti-S1 antibody and the cDNA are useful for diagnosing,  
CC preventing, treating or ameliorating a medical condition in mammalian  
CC subject especially diseases and/or disorders of the ovary  
CC and/or breast such as neoplastic disorders (such as ovarian Krukenberg  
CC tumour and cancer), infectious diseases (e.g., mastitis, oophoritis),  
CC inflammatory diseases (e.g., abscesses), reproductive system disorders  
CC (Paget's disease), autoimmune disorders (systemic lupus erythematosus,  
CC rheumatoid arthritis), blood-related disorders (sickle cell anaemia),  
CC hyperproliferative disorders, urinary system disorders  
CC (glomerulonephritis), cardiovascular disorders (arrhythmias),  
CC respiratory disorders, musculoskeletal system disorders, neural  
CC activity and neurological disorders (Alzheimer's disease and  
CC Parkinson's disease), endocrine disorders (Addison's disease),  
CC gastrointestinal disorders (inflammatory disorders), liver disorders  
CC (biliary liver cirrhosis), pancreatic and gall bladder disorders,  
CC disorders of the large intestine, developmental and inherited  
CC disorders, diseases at the cellular level, and wound healing and  
CC epithelial cell proliferation. They are also useful to prevent skin  
CC aging, for preventing hair loss, to maintain organs before  
CC transplantation or for supporting cell culture of primary tissues, to  
CC modulate mammalian characteristics such as body height, to modulate  
CC mammalian metabolism, to change a mammal's mental or physical state,  
CC and as food additive or preservative. The present sequence  
CC represents an ovarian antigen, S1 protein of the invention.  
CC Note: The sequence data for this patent did not form part  
CC of the printed specification, but was obtained in electronic  
CC format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences.  
CC Note: The sequence data for this patent did not form part  
CC of the printed specification, but was obtained in electronic  
CC format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences.

XX

SQ Sequence 56 AA;

Query Match 26.6%; Score 51; DB 22; Length 56;  
Best Local Similarity 57.9%; Pred. No. 18;  
Matches 11; Conservative 1; Mismatches 5; Indels 2; Gaps 1;

Qy 9 LNCWMLSAFSRYAR--CLA 25

|||| || |: | |||

Db 25 LNCWHLSCFNHALRLSCLA 43

RESULT 10

AAM94423

ID AAM94423 standard; Protein; 56 AA.

XX

AC AAM94423;

XX

DT 21-NOV-2001 (first entry)

XX

DE Human reproductive system related antigen SEQ ID NO: 3081.

XX

KW Human; reproductive system related antigen; reproductive system disorder;  
KW cancer; gene therapy.

XX

OS Homo sapiens.

XX

PN WO200155320-A2.

XX

PD 02-AUG-2001.

XX

PF 17-JAN-2001; 2001WO-US01339.

XX

PR 31-JAN-2000; 2000US-0179065.

PR 04-FEB-2000; 2000US-0180628.

PR 24-FEB-2000; 2000US-0184664.

PR 02-MAR-2000; 2000US-0186350.

PR 16-MAR-2000; 2000US-0189874.

PR 17-MAR-2000; 2000US-0190076.

PR 18-APR-2000; 2000US-0198123.

PR 19-MAY-2000; 2000US-0205515.

PR 07-JUN-2000; 2000US-0209467.

PR 28-JUN-2000; 2000US-0214886.

PR 30-JUN-2000; 2000US-0215135.

PR 07-JUL-2000; 2000US-0216647.

PR 07-JUL-2000; 2000US-0216880.

PR 11-JUL-2000; 2000US-0217487.

PR 11-JUL-2000; 2000US-0217496.

PR 14-JUL-2000; 2000US-0218290.

PR 26-JUL-2000; 2000US-0220963.

PR 26-JUL-2000; 2000US-0220964.

PR 14-AUG-2000; 2000US-0224518.

PR 14-AUG-2000; 2000US-0224519.

PR 14-AUG-2000; 2000US-0225213.

PR 14-AUG-2000; 2000US-0225214.

PR 14-AUG-2000; 2000US-0225266.

PR 14-AUG-2000; 2000US-0225267.

PR 14-AUG-2000; 2000US-0225268.

PR 14-AUG-2000; 2000US-0225270.

PR 14-AUG-2000; 2000US-0225447.

PR 14-AUG-2000; 2000US-0225757.

PR 14-AUG-2000; 2000US-0225758.

PR 14-AUG-2000; 2000US-0225759.

PR 18-AUG-2000; 2000US-0226279.

PR 22-AUG-2000; 2000US-0226681.

PR 22-AUG-2000; 2000US-0226868.

PR 22-AUG-2000; 2000US-0227182.

PR 23-AUG-2000; 2000US-0227009.

PR 30-AUG-2000; 2000US-0228924.

PR 01-SEP-2000; 2000US-0229287.  
PR 01-SEP-2000; 2000US-0229343.  
PR 01-SEP-2000; 2000US-0229344.  
PR 01-SEP-2000; 2000US-0229345.  
PR 05-SEP-2000; 2000US-0229509.  
PR 05-SEP-2000; 2000US-0229513.  
PR 06-SEP-2000; 2000US-0230437.  
PR 06-SEP-2000; 2000US-0230438.  
PR 08-SEP-2000; 2000US-0231242.  
PR 08-SEP-2000; 2000US-0231243.  
PR 08-SEP-2000; 2000US-0231244.  
PR 08-SEP-2000; 2000US-0231413.  
PR 08-SEP-2000; 2000US-0231414.  
PR 08-SEP-2000; 2000US-0232080.  
PR 08-SEP-2000; 2000US-0232081.  
PR 12-SEP-2000; 2000US-0231968.  
PR 14-SEP-2000; 2000US-0232397.  
PR 14-SEP-2000; 2000US-0232398.  
PR 14-SEP-2000; 2000US-0232399.  
PR 14-SEP-2000; 2000US-0232400.  
PR 14-SEP-2000; 2000US-0232401.  
PR 14-SEP-2000; 2000US-0233063.  
PR 14-SEP-2000; 2000US-0233064.  
PR 14-SEP-2000; 2000US-0233065.  
PR 21-SEP-2000; 2000US-0234223.  
PR 21-SEP-2000; 2000US-0234274.  
PR 25-SEP-2000; 2000US-0234997.  
PR 25-SEP-2000; 2000US-0234998.  
PR 26-SEP-2000; 2000US-0235484.  
PR 27-SEP-2000; 2000US-0235834.  
PR 27-SEP-2000; 2000US-0235836.  
PR 29-SEP-2000; 2000US-0236327.  
PR 29-SEP-2000; 2000US-0236367.  
PR 29-SEP-2000; 2000US-0236368.  
PR 29-SEP-2000; 2000US-0236369.  
PR 29-SEP-2000; 2000US-0236370.  
PR 02-OCT-2000; 2000US-0236802.  
PR 02-OCT-2000; 2000US-0237037.  
PR 02-OCT-2000; 2000US-0237038.  
PR 02-OCT-2000; 2000US-0237039.  
PR 02-OCT-2000; 2000US-0237040.  
PR 13-OCT-2000; 2000US-0239935.  
PR 13-OCT-2000; 2000US-0239937.  
PR 20-OCT-2000; 2000US-0240960.  
PR 20-OCT-2000; 2000US-0241221.  
PR 20-OCT-2000; 2000US-0241785.  
PR 20-OCT-2000; 2000US-0241786.  
PR 20-OCT-2000; 2000US-0241787.  
PR 20-OCT-2000; 2000US-0241808.  
PR 20-OCT-2000; 2000US-0241809.  
PR 20-OCT-2000; 2000US-0241826.  
PR 01-NOV-2000; 2000US-0244617.  
PR 08-NOV-2000; 2000US-0246474.  
PR 08-NOV-2000; 2000US-0246475.  
PR 08-NOV-2000; 2000US-0246476.  
PR 08-NOV-2000; 2000US-0246477.  
PR 08-NOV-2000; 2000US-0246478.

PR 08-NOV-2000; 2000US-0246523.  
PR 08-NOV-2000; 2000US-0246524.  
PR 08-NOV-2000; 2000US-0246525.  
PR 08-NOV-2000; 2000US-0246526.  
PR 08-NOV-2000; 2000US-0246527.  
PR 08-NOV-2000; 2000US-0246528.  
PR 08-NOV-2000; 2000US-0246532.  
PR 08-NOV-2000; 2000US-0246609.  
PR 08-NOV-2000; 2000US-0246610.  
PR 08-NOV-2000; 2000US-0246611.  
PR 08-NOV-2000; 2000US-0246613.  
PR 17-NOV-2000; 2000US-0249207.  
PR 17-NOV-2000; 2000US-0249208.  
PR 17-NOV-2000; 2000US-0249209.  
PR 17-NOV-2000; 2000US-0249210.  
PR 17-NOV-2000; 2000US-0249211.  
PR 17-NOV-2000; 2000US-0249212.  
PR 17-NOV-2000; 2000US-0249213.  
PR 17-NOV-2000; 2000US-0249214.  
PR 17-NOV-2000; 2000US-0249215.  
PR 17-NOV-2000; 2000US-0249216.  
PR 17-NOV-2000; 2000US-0249217.  
PR 17-NOV-2000; 2000US-0249218.  
PR 17-NOV-2000; 2000US-0249244.  
PR 17-NOV-2000; 2000US-0249245.  
PR 17-NOV-2000; 2000US-0249264.  
PR 17-NOV-2000; 2000US-0249265.  
PR 17-NOV-2000; 2000US-0249297.  
PR 17-NOV-2000; 2000US-0249299.  
PR 17-NOV-2000; 2000US-0249300.  
PR 01-DEC-2000; 2000US-0250160.  
PR 01-DEC-2000; 2000US-0250391.  
PR 05-DEC-2000; 2000US-0251030.  
PR 05-DEC-2000; 2000US-0251988.  
PR 05-DEC-2000; 2000US-0256719.  
PR 06-DEC-2000; 2000US-0251479.  
PR 08-DEC-2000; 2000US-0251856.  
PR 08-DEC-2000; 2000US-0251868.  
PR 08-DEC-2000; 2000US-0251869.  
PR 08-DEC-2000; 2000US-0251989.  
PR 08-DEC-2000; 2000US-0251990.  
PR 11-DEC-2000; 2000US-0254097.  
PR 05-JAN-2001; 2001US-0259678.

XX

PA (HUMA-) HUMAN GENOME SCI INC.

XX

PI Rosen CA, Barash SC, Ruben SM;

XX

DR WPI; 2001-465570/50.

DR N-PSDB; AAL00393.

XX

PT Isolated nucleic acid molecule encoding a reproductive system antigen  
PT is used in preventing, treating or ameliorating a medical condition -

XX

PS Claim 11; SEQ ID NO 3081; 1297pp + Sequence Listing; English.

XX

CC The present invention provides the protein and coding sequences of a



CC number of human reproductive system related antigens. These can be used  
CC in the prevention and treatment of reproductive system disorders,  
CC including cancer. The present sequence is a protein of the invention.

XX

SQ Sequence 56 AA;

Query Match 26.6%; Score 51; DB 22; Length 56;  
Best Local Similarity 57.9%; Pred. No. 18;  
Matches 11; Conservative 1; Mismatches 5; Indels 2; Gaps 1;

QY 9 LNCWMLSAFSRYAR--CLA 25  
||| | | : | |||  
Db 25 LNCWHLSCFNHALRLSCLA 43

# RESULT 11

ABG61724

ID ABG61724 standard; Protein; 56 AA.

XX

AC ABG61724;

XX

DT 26-AUG-2002 (first entry)

XX

DE Novel ovarian related polypeptide #15.

XX

KW Ovarian related polypeptide; neoplastic disorder; tumour; ovarian cancer;  
KW hyperproliferative disorder; adult acute lymphocytic leukaemia;  
KW breast cancer; reproductive system disorder; tuberculosis; arthritis;  
KW immune system disorder; Chediak-Higashi's syndrome; neonatal neutropenia;  
KW autoimmune disorder; Hashimoto's thyroiditis; inflammatory disorder;  
KW septic shock; multiple sclerosis; central nervous system disorder;  
KW neurological disorder; allergy; Parkinson's disease; Alzheimer's disease;  
KW cardiovascular disorder; atherosclerosis; blood related disorder;  
KW respiratory disorder; urinary system disorder; musculoskeletal disorder;  
KW osteoporosis; wound healing; endocrine disorder; infectious disease;  
KW gastrointestinal disorder; transplantation; food additive; preservative.

XX

OS Homo sapiens.

XX

PN US2002045230-A1.

XX

PD 18-APR-2002.

XX

PF 20-JUL-2001; 2001US-0908711.

XX

PR 31-JAN-2000; 2000US-179065P.

PR 04-FEB-2000; 2000US-180628P.

PR 24-FEB-2000; 2000US-184664P.

PR 02-MAR-2000; 2000US-186350P.

PR 16-MAR-2000; 2000US-189874P.

PR 17-MAR-2000; 2000US-190076P.

PR 18-APR-2000; 2000US-198123P.

PR 19-MAY-2000; 2000US-205515P.

PR 07-JUN-2000; 2000US-209467P.

PR 28-JUN-2000; 2000US-214886P.

PR 30-JUN-2000; 2000US-215135P.

PR 07-JUL-2000; 2000US-216647P.

PR 07-JUL-2000; 2000US-216880P.  
PR 11-JUL-2000; 2000US-217487P.  
PR 11-JUL-2000; 2000US-217496P.  
PR 14-JUL-2000; 2000US-218290P.  
PR 26-JUL-2000; 2000US-220963P.  
PR 26-JUL-2000; 2000US-220964P.  
PR 14-AUG-2000; 2000US-224518P.  
PR 14-AUG-2000; 2000US-224519P.  
PR 14-AUG-2000; 2000US-225213P.  
PR 14-AUG-2000; 2000US-225214P.  
PR 14-AUG-2000; 2000US-225266P.  
PR 14-AUG-2000; 2000US-225267P.  
PR 14-AUG-2000; 2000US-225268P.  
PR 14-AUG-2000; 2000US-225270P.  
PR 14-AUG-2000; 2000US-225447P.  
PR 14-AUG-2000; 2000US-225757P.  
PR 14-AUG-2000; 2000US-225758P.  
PR 14-AUG-2000; 2000US-225759P.  
PR 18-AUG-2000; 2000US-226279P.  
PR 22-AUG-2000; 2000US-226681P.  
PR 22-AUG-2000; 2000US-226868P.  
PR 22-AUG-2000; 2000US-227182P.  
PR 23-AUG-2000; 2000US-227009P.  
PR 30-AUG-2000; 2000US-228924P.  
PR 01-SEP-2000; 2000US-229287P.  
PR 01-SEP-2000; 2000US-229343P.  
PR 01-SEP-2000; 2000US-229344P.  
PR 01-SEP-2000; 2000US-229345P.  
PR 05-SEP-2000; 2000US-229509P.  
PR 05-SEP-2000; 2000US-229513P.  
PR 06-SEP-2000; 2000US-230437P.  
PR 06-SEP-2000; 2000US-230438P.  
PR 08-SEP-2000; 2000US-231242P.  
PR 08-SEP-2000; 2000US-231243P.  
PR 08-SEP-2000; 2000US-231244P.  
PR 08-SEP-2000; 2000US-231413P.  
PR 08-SEP-2000; 2000US-231414P.  
PR 08-SEP-2000; 2000US-232080P.  
PR 08-SEP-2000; 2000US-232081P.  
PR 12-SEP-2000; 2000US-231968P.  
PR 14-SEP-2000; 2000US-232397P.  
PR 14-SEP-2000; 2000US-232398P.  
PR 14-SEP-2000; 2000US-232399P.  
PR 14-SEP-2000; 2000US-232400P.  
PR 14-SEP-2000; 2000US-232401P.  
PR 14-SEP-2000; 2000US-233063P.  
PR 14-SEP-2000; 2000US-233064P.  
PR 14-SEP-2000; 2000US-233065P.  
PR 21-SEP-2000; 2000US-234223P.  
PR 21-SEP-2000; 2000US-234274P.  
PR 25-SEP-2000; 2000US-234997P.  
PR 25-SEP-2000; 2000US-234998P.  
PR 26-SEP-2000; 2000US-235484P.  
PR 27-SEP-2000; 2000US-235834P.  
PR 27-SEP-2000; 2000US-235836P.  
PR 29-SEP-2000; 2000US-236327P.  
PR 29-SEP-2000; 2000US-236367P.

PR 29-SEP-2000; 2000US-236368P.  
PR 29-SEP-2000; 2000US-236369P.  
PR 29-SEP-2000; 2000US-236370P.  
PR 02-OCT-2000; 2000US-236802P.  
PR 02-OCT-2000; 2000US-237037P.  
PR 02-OCT-2000; 2000US-237038P.  
PR 02-OCT-2000; 2000US-237039P.  
PR 02-OCT-2000; 2000US-237040P.  
PR 13-OCT-2000; 2000US-239935P.  
PR 13-OCT-2000; 2000US-239937P.  
PR 20-OCT-2000; 2000US-240960P.  
PR 20-OCT-2000; 2000US-241221P.  
PR 20-OCT-2000; 2000US-241785P.  
PR 20-OCT-2000; 2000US-241786P.  
PR 20-OCT-2000; 2000US-241787P.  
PR 20-OCT-2000; 2000US-241808P.  
PR 20-OCT-2000; 2000US-241809P.  
PR 20-OCT-2000; 2000US-241826P.  
PR 01-NOV-2000; 2000US-244617P.  
PR 08-NOV-2000; 2000US-246474P.  
PR 08-NOV-2000; 2000US-246475P.  
PR 08-NOV-2000; 2000US-246476P.  
PR 08-NOV-2000; 2000US-246477P.  
PR 08-NOV-2000; 2000US-246478P.  
PR 08-NOV-2000; 2000US-246523P.  
PR 08-NOV-2000; 2000US-246524P.  
PR 08-NOV-2000; 2000US-246525P.  
PR 08-NOV-2000; 2000US-246526P.  
PR 08-NOV-2000; 2000US-246527P.  
PR 08-NOV-2000; 2000US-246528P.  
PR 08-NOV-2000; 2000US-246532P.  
PR 08-NOV-2000; 2000US-246609P.  
PR 08-NOV-2000; 2000US-246610P.  
PR 08-NOV-2000; 2000US-246611P.  
PR 08-NOV-2000; 2000US-246613P.  
PR 17-NOV-2000; 2000US-249207P.  
PR 17-NOV-2000; 2000US-249208P.  
PR 17-NOV-2000; 2000US-249209P.  
PR 17-NOV-2000; 2000US-249210P.  
PR 17-NOV-2000; 2000US-249211P.  
PR 17-NOV-2000; 2000US-249212P.  
PR 17-NOV-2000; 2000US-249213P.  
PR 17-NOV-2000; 2000US-249214P.  
PR 17-NOV-2000; 2000US-249215P.  
PR 17-NOV-2000; 2000US-249216P.  
PR 17-NOV-2000; 2000US-249217P.  
PR 17-NOV-2000; 2000US-249218P.  
PR 17-NOV-2000; 2000US-249244P.  
PR 17-NOV-2000; 2000US-249245P.  
PR 17-NOV-2000; 2000US-249264P.  
PR 17-NOV-2000; 2000US-249265P.  
PR 17-NOV-2000; 2000US-249297P.  
PR 17-NOV-2000; 2000US-249299P.  
PR 17-NOV-2000; 2000US-249300P.  
PR 01-DEC-2000; 2000US-250160P.  
PR 01-DEC-2000; 2000US-250391P.  
PR 05-DEC-2000; 2000US-251030P.

PR 05-DEC-2000; 2000US-251988P.  
 PR 05-DEC-2000; 2000US-256719P.  
 PR 06-DEC-2000; 2000US-251479P.  
 PR 08-DEC-2000; 2000US-251856P.  
 PR 08-DEC-2000; 2000US-251868P.  
 PR 08-DEC-2000; 2000US-251869P.  
 PR 08-DEC-2000; 2000US-251989P.  
 PR 08-DEC-2000; 2000US-251990P.  
 PR 11-DEC-2000; 2000US-254097P.  
 PR 05-JAN-2001; 2001US-259678P.  
 PR 17-JAN-2001; 2001US-0764853.  
 PR 17-JAN-2001; 2001US-0764856.  
 PR 17-JAN-2001; 2001US-0764864.  
 PR 17-JAN-2001; 2001US-0764867.  
 PR 17-JAN-2001; 2001US-0764868.  
 PR 17-JAN-2001; 2001US-0764869.  
 PR 17-JAN-2001; 2001US-0764870.  
 PR 17-JAN-2001; 2001US-0764874.  
 PR 17-JAN-2001; 2001US-0764882.  
 PR 17-JAN-2001; 2001US-0764888.  
 PR 17-JAN-2001; 2001US-0764891.  
 PR 17-JAN-2001; 2001US-0764892.  
 PR 17-JAN-2001; 2001US-0764896.  
 PR 17-JAN-2001; 2001US-0764898.  
 PR 17-JAN-2001; 2001US-0764902.  
 PR 17-JAN-2001; 2001US-0764905.  
 PR 17-JAN-2001; 2001WO-US01239.  
 PR 17-JAN-2001; 2001WO-US01307.  
 PR 17-JAN-2001; 2001WO-US01312.  
 PR 17-JAN-2001; 2001WO-US01320.  
 PR 17-JAN-2001; 2001WO-US01329.  
 PR 17-JAN-2001; 2001WO-US01334.  
 PR 17-JAN-2001; 2001WO-US01336.  
 PR 17-JAN-2001; 2001WO-US01339.  
 PR 17-JAN-2001; 2001WO-US01340.  
 PR 17-JAN-2001; 2001WO-US01341.  
 PR 17-JAN-2001; 2001WO-US01344.  
 PR 17-JAN-2001; 2001WO-US01345.  
 PR 17-JAN-2001; 2001WO-US01347.  
 PR 17-JAN-2001; 2001WO-US01348.  
 PR 17-JAN-2001; 2001WO-US01360.

XX

PA (ROSE/) ROSEN C A.  
 PA (RUBI/) RUBIN S M.  
 PA (BARA/) BARASH S C.

XX

Query Match 26.6%; Score 51; DB 23; Length 56;  
 Best Local Similarity 57.9%; Pred. No. 18;  
 Matches 11; Conservative 1; Mismatches 5; Indels 2; Gaps 1;

Qy 9 LNCWMLSAFSRYAR--CLA 25  
 |||| || |: | |||  
 Db 25 LNCWHLSCFNHALRLSCIA 43

RESULT 12

AAE23630

ID AAE23630 standard; Protein; 466 AA.

XX

AC AAE23630;

XX

DT 27-AUG-2002 (first entry)

XX

DE Escherichia coli 6-phospho-strand-glucosidase #1.

XX

KW Secondary structural element; genome-sized database; EC 3.2.1.86;

KW enzyme; 6-phospho-strand-glucosidase.

XX

OS Escherichia coli.

XX

FH Key Location/Qualifiers

FT Region 9..11

FT /note= "Strand A"

FT Region 12..103

FT /note= "GG dipeptide parse"

FT Region 14..20

FT /note= "Strand B"

FT Region 48..59

FT /note= "Parse region"

FT Region 49..51

FT /note= "Strand A"

FT Region 50..54

FT /note= "Parsing string (PGDSG)"

FT Region 53..60

FT /note= "Strand B"

FT Region 62..63

FT /note= "GG dipeptide parse"

FT Region 71..81

FT /note= "Parse region"

FT Region 72..84

FT /note= "Helix 1"

FT Region 73..78

FT /note= "Beta strand"

FT Region 89..93

FT /note= "Strand C"

FT Region 95..107

FT /note= "Helix 1"

FT Active-site 95..102

FT /note= "Active site a"

FT Region 111..115

FT /note= "Strand C"

FT Region 116..130

FT /note= "Helix 2"

FT Active-site 117..125

FT /note= "Active site a"

FT Region 136..140

FT /note= "Strand D"

FT Region 138..153

FT /note= "Helix 2"

FT Active-site 141

FT /note= "Active site b"

FT Region 158..177

FT /note= "Helix 3"

FT	Region	159..163
FT		/note= "Strand D"
FT	Active-site	164..166
FT		/note= "Active site b"
FT	Region	181..198
FT		/note= "Helix 3"
FT	Region	182..185
FT		/note= "Strand E"
FT	Active-site	184..187
FT		/note= "Active site c"
FT	Active-site	205..209
FT		/note= "Active site c"
FT	Region	205..208
FT		/note= "Strand E"
FT	Active-site	207..209
FT		/note= "Active site c"
FT	Region	212..215
FT		/note= "Beta strand"
FT	Region	212..226
FT		/note= "Helix 4"
FT	Region	219..221
FT		/note= "Beta strand"
FT	Region	242..245
FT		/note= "Strand X"
FT	Region	248..268
FT		/note= "Helix 4"
FT	Region	259..273
FT		/note= "Helix X"
FT	Region	275..280
FT		/note= "Strand Y"
FT	Region	286..293
FT		/note= "Helix 5"
FT	Region	318..320
FT		/note= "Strand Y"
FT	Region	323..327
FT		/note= "Strand F"
FT	Region	329..339
FT		/note= "Helix Y"
FT	Region	332..342
FT		/note= "Helix 5"
FT	Region	375..382
FT		/note= "Strand Z"
FT	Region	381..388
FT		/note= "Strand F"
FT	Region	385..398
FT		/note= "Helix 6"
FT	Region	404..407
FT		/note= "Strand G"
FT	Active-site	408..410
FT		/note= "Active site d"
FT	Region	431..448
FT		/note= "Helix 7"
FT	Region	446..452
FT		/note= "Strand Z"
FT	Region	450..454
FT		/note= "Strand H"
FT	Region	456..459

FT /note= "Strand I"  
 FT Region 456..469  
 FT /note= "Helix 6"  
 FT Region 460..462  
 FT /note= "Dipeptide GP parse"  
 FT Region 464..467  
 FT /note= "Strand J"  
 FT Region 476..479  
 FT /note= "Strand G"  
 FT Region 478..482  
 FT /note= "Strand K"  
 FT Active-site 480..482  
 FT /note= "Active site d"  
 FT Region 496..509  
 FT /note= "Helix 8"  
 FT Region 497..517  
 FT /note= "Helix 7"  
 FT Region 521..525  
 FT /note= "Strand H"  
 FT Region 527..530  
 FT /note= "Strand I"  
 FT Region 535..539  
 FT /note= "Strand J"  
 FT Region 548..554  
 FT /note= "Strand K"  
 FT Region 563..576  
 FT /note= "Helix 8"  
 XX  
 PN US6377893-B1.  
 XX  
 PD 23-APR-2002.  
 XX  
 PF 19-AUG-1997; 97US-0914375.  
 XX  
 PR 25-MAR-1992; 92US-0857224.  
 XX  
 PA (BENN/) BENNER S A.  
 XX  
 PI Benner SA;  
 XX  
 DR WPI; 2002-424771/45.  
 XX  
 PT Methods for excluding or detecting homology between protein families,  
 PT useful e.g. for identifying in vitro properties of proteins important  
 PT for physiological activity -  
 XX  
 PS Example 5; Column 147-150; 99pp; English.  
 XX  
 CC The invention relates to a method for excluding homology between  
 CC two protein families. The method involves constructing models for  
 CC secondary structural elements for each family; aligning secondary  
 CC structural elements of one family with the secondary structural  
 CC elements from the other family around sequence motifs; determining  
 CC whether secondary structural elements flanking the sequence motifs  
 CC in one family are congruent to secondary structural elements in  
 CC the other family, so as to determine if the families are related  
 CC by common ancestry or not. The method is used to confirm/deny the

CC hypothesis that proteins are homologous and related methods are  
CC used to identify mutations during divergent evolution of proteins,  
CC to identify in vitro properties of proteins that are important for  
CC physiological activity and to generate genome-sized databases.  
CC The present sequence is Escherichia coli 6-phospho-strand-glucosidase  
CC (EC 3.2.1.86). This sequence is used in the exemplification of the  
CC invention.

XX

SQ Sequence 466 AA;

Query Match 26.6%; Score 51; DB 23; Length 466;  
Best Local Similarity 43.3%; Pred. No. 1.8e+02;  
Matches 13; Conservative 3; Mismatches 14; Indels 0; Gaps 0;

QY 1 LGTFWGD'TLNCWMLSAFSRYARCLAEGHDG 30  
| | :| | :| | | | | | | |  
Db 138 LVTEYGSWRNRKLVEFFSRYARTCFEAFDG 167

RESULT 13

AAU34677

ID AAU34677 standard; Protein; 474 AA.

XX

AC AAU34677;

XX

DT 14-FEB-2002 (first entry)

XX

DE E. coli cellular proliferation protein #258.

XX

KW Antisense; prokaryotic cellular proliferation protein;

KW antibiotic; antibacterial; drug design.

XX

OS Escherichia coli.

XX

PN WO200170955-A2.

XX

PD 27-SEP-2001.

XX

PF 21-MAR-2001; 2001WO-US09180.

XX

PR 21-MAR-2000; 2000US-191078P.

PR 23-MAY-2000; 2000US-206848P.

PR 26-MAY-2000; 2000US-207727P.

PR 23-OCT-2000; 2000US-242578P.

PR 27-NOV-2000; 2000US-253625P.

PR 22-DEC-2000; 2000US-257931P.

PR 16-FEB-2001; 2001US-269308P.

XX

PA (ELIT-) ELITRA PHARM INC.

XX

PI Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;

PI Yamamoto RT, Xu HH;

XX

DR WPI; 2001-611495/70.

DR N-PSDB; AAS52536.

XX

PT New polynucleotides for the identification and development of



PT antibiotics, comprise sequences of antisense nucleic acids -  
 XX  
 PS Example 3; Seq ID No 10270; 511pp; English.  
 XX  
 CC The invention relates to antisense inhibitors of genes essential to  
 CC prokaryotic cellular proliferation, their use in identifying the  
 CC genes, their use in the discovery of novel antibiotics, the essential  
 CC genes themselves and the encoded proteins. The prokaryotes used are  
 CC Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella  
 CC pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The  
 CC invention is also useful for the identification of potential new targets  
 CC for antibiotic development. The antisense nucleic acids can also be used  
 CC to identify proteins used in proliferation, to express these proteins,  
 CC and to obtain antibodies capable of binding to the expressed proteins.  
 CC The proteins can be used to screen compounds in rational drug discovery  
 CC programmes. The antisense nucleic acid sequence is also useful to screen  
 CC for homologous nucleic acids which are required for cell proliferation in  
 CC a wide variety of organisms. The present sequence represents an  
 CC essential prokaryotic cellular proliferation protein.  
 CC Note: The sequence data for this patent did not form part  
 CC of the printed specification, but was obtained in electronic  
 CC format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences.  
 XX  
 SQ Sequence 474 AA;

Query Match 26.6%; Score 51; DB 22; Length 474;  
 Best Local Similarity 43.3%; Pred. No. 1.8e+02;  
 Matches 13; Conservative 3; Mismatches 14; Indels 0; Gaps 0;

QY 1 LGTFWGDTLNCWMLSAFSRYARCLAEGHDG 30  
 | | : | | : | | | | | | |  
 Db 141 LVTEYGSWRNRLVEFFSRYARTCFEAFDG 170

# RESULT 14

AAM78524

ID AAM78524 standard; Protein; 1207 AA.

XX

AC AAM78524;

XX

DT 06-NOV-2001 (first entry)

XX

DE Human protein SEQ ID NO 1186.

XX

KW Human; cytokine; cell proliferation; cell differentiation; gene therapy;  
 KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;  
 KW tissue growth factor; immunomodulatory; cancer; leukaemia;  
 KW nervous system disorder; arthritis; inflammation.

XX

OS Homo sapiens.

XX

PN WO200157190-A2.

XX

PD 09-AUG-2001.

XX

PF 05-FEB-2001; 2001WO-US04098.

XX  
PR 03-FEB-2000; 2000US-0496914.  
PR 27-APR-2000; 2000US-0560875.  
PR 20-JUN-2000; 2000US-0598075.  
PR 19-JUL-2000; 2000US-0620325.  
PR 01-SEP-2000; 2000US-0654936.  
PR 15-SEP-2000; 2000US-0663561.  
PR 20-OCT-2000; 2000US-0693325.  
PR 30-NOV-2000; 2000US-0728422.  
XX  
PA (HYSE-) HYSEQ INC.  
XX  
PI Tang YT, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y, Ma Y;  
PI Zhao QA, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW;  
PI Xue AJ, Yang Y, Wejhrman T, Goodrich R;  
XX  
DR WPI; 2001-476283/51.  
DR N-PSDB; AAK51657.  
XX  
PT Nucleic acids encoding polypeptides with cytokine-like activities,  
PT useful in diagnosis and gene therapy -  
XX  
PS Claim 20; Page 3434-3436; 6221pp; English.  
XX  
CC The invention relates to polynucleotides (AAK51456-AAK53435) and the  
CC encoded polypeptides (AAM78323-AAM80302) that exhibit activity elating to  
CC cytokine, cell proliferation or cell differentiation or which may induce  
CC production of other cytokines in other cell populations. The  
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or  
CC peptide therapy. The polypeptides have various cytokine-like activities,  
CC e.g. stem cell growth factor activity, haematopoiesis regulating  
CC activity, tissue growth factor activity, immunomodulatory activity and  
CC activin/inhibin activity and may be useful in the diagnosis and/or  
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and  
CC inflammation.  
CC Note: Records for SEQ ID NO 2110 (AAK52581), 2111 (AAK52582) and 3666  
CC (AAM80020) are omitted as the relevant pages from the sequence listing  
CC were missing at the time of publication.  
XX  
SQ Sequence 1207 AA;

Query Match 26.6%; Score 51; DB 22; Length 1207;  
Best Local Similarity 56.2%; Pred. No. 4.9e+02;  
Matches 9; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 18 SRYARCLAEAGHDGPTQ 33  
| ||||::|| | |  
Db 841 SMYARCISEGEDATCQ 856

RESULT 15

AAB84604

ID AAB84604 standard; Protein; 1207 AA.

XX

AC AAB84604;

XX

DT 05-SEP-2001 (first entry)

XX  
 DE Amino acid sequence of endothelial growth factor.  
 XX  
 KW Growth factor; protein inhibitor; protease; damaged tissue;  
 KW platelet-derived growth factor; PDGF; fibroblast growth factor; FGF;  
 KW connective tissue derived growth factor; CTGF; chrysalin; VEGF;  
 KW keratinocyte-derived growth factor; KGF; epidermal growth factor; EGF;  
 KW transforming growth factor-beta; TGF-beta; matrix metalloproteinase; MMP;  
 KW granulocyte macrophage colony stimulating factor; GM-CSF; uPA;  
 KW vascular endothelial growth factor; urokinase plasminogen activator;  
 KW dermal ulcer; wound.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200149309-A2.  
 XX  
 PD 12-JUL-2001.  
 XX  
 PF 21-DEC-2000; 2000WO-IB01935.  
 XX  
 PR 29-DEC-1999; 99GB-0030768.  
 XX  
 PA (PFIZ ) PFIZER LTD.  
 PA (PFIZ ) PFIZER INC.  
 XX  
 PI Davies MJ, Huggins JP, McIntosh FS, Occleston NL;  
 XX  
 DR WPI; 2001-418351/44.  
 DR N-PSDB; AAH28219.  
 XX  
 PT Composition for the treatment of damaged tissue i.e. chronic wounds and  
 PT dermal ulcers comprises an inhibitor agent i.e. a protease and a growth  
 PT factor -  
 XX  
 PS Disclosure; Page 549; 572pp; English.  
 XX  
 CC The specification describes a pharmaceutical composition, comprising  
 CC a growth factor, an inhibitor agent, i.e. a protease. The inhibitor  
 CC agent inhibits the action of at least one specific adverse protein,  
 CC i.e. a protease, that is upregulated in a damaged tissue such as a  
 CC wound environment. Growth factors which are included in the composition  
 CC of the invention are platelet-derived growth factor (PDGF), fibroblast  
 CC growth factor (FGF), connective tissue derived growth factor (CTGF),  
 CC keratinocyte-derived growth factor (KGF), transforming growth  
 CC factor-beta (TGF-beta), granulocyte macrophage colony stimulating factor  
 CC (GM-CSF), epidermal growth factor (EGF), vascular endothelial growth  
 CC factor (VEGF), and chrysalin. Inhibitors which are included in the  
 CC composition of the invention include inhibitors of urokinase-type  
 CC plasminogen activator (uPA) and matrix metalloproteinase (MMP). The  
 CC composition is useful for the treatment of chronic damaged tissue, i.e.  
 CC wounds and dermal ulcers. The present sequence represents a human EGF,  
 CC and is used to produce the composition of the invention.  
 XX  
 SQ Sequence 1207 AA;

Query Match 26.6%; Score 51; DB 22; Length 1207;  
 Best Local Similarity 56.2%; Pred. No. 4.9e+02;

Matches 9; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

Qy 18 SRYARCLAE~~G~~HDGPTQ 33  
| ||||::|| | |  
Db 841 SMYARCISEGEDATCQ 856

Search completed: January 30, 2004, 11:24:36  
Job time : 5.87938 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 30, 2004, 11:23:12 ; Search time 1.86187 Seconds  
(without alignments)  
749.923 Million cell updates/sec

Title: US-09-989-481-4  
Perfect score: 192  
Sequence: 1 LGTFWGDTLNCWMLSAFSRYARCLAEGHDGPTQ 33

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued\_Patents\_AA:\*  
1: /cgn2\_6/ptodata/1/iaa/5A\_COMB.pep:\*  
2: /cgn2\_6/ptodata/1/iaa/5B\_COMB.pep:\*  
3: /cgn2\_6/ptodata/1/iaa/6A\_COMB.pep:\*  
4: /cgn2\_6/ptodata/1/iaa/6B\_COMB.pep:\*  
5: /cgn2\_6/ptodata/1/iaa/PCTUS\_COMB.pep:\*  
6: /cgn2\_6/ptodata/1/iaa/backfiles1.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query		DB	ID	Description
		Match	Length			
1	192	100.0	33	4	US-08-928-862-4	Sequence 4, Appli
2	52	27.1	385	4	US-09-252-991A-19628	Sequence 19628, A
3	51	26.6	466	4	US-08-914-375C-60	Sequence 60, Appl
4	49	25.5	341	4	US-09-328-352-6124	Sequence 6124, Ap
5	48	25.0	113	3	US-08-836-075A-80	Sequence 80, Appl
6	48	25.0	1940	2	US-08-644-271-30	Sequence 30, Appl
7	48	25.0	1940	4	US-09-077-955-34	Sequence 34, Appl
8	47.5	24.7	477	4	US-09-252-991A-17047	Sequence 17047, A
9	47.5	24.7	865	4	US-09-612-204B-24	Sequence 24, Appl
10	47	24.5	230	4	US-09-252-991A-30553	Sequence 30553, A
11	47	24.5	736	4	US-09-252-991A-22859	Sequence 22859, A

12	47	24.5	1778	4	US-09-252-991A-18159	Sequence 18159, A
13	46.5	24.2	243	3	US-09-191-647-14	Sequence 14, Appl
14	46.5	24.2	243	3	US-09-540-245A-14	Sequence 14, Appl
15	46.5	24.2	243	3	US-09-540-153-14	Sequence 14, Appl
16	46.5	24.2	716	4	US-09-312-283C-183	Sequence 183, App
17	46.5	24.2	771	3	US-09-188-930-183	Sequence 183, Appl
18	46.5	24.2	1525	3	US-09-191-647-2	Sequence 2, Appli
19	46.5	24.2	1525	3	US-09-540-245A-2	Sequence 2, Appli
20	46.5	24.2	1525	3	US-09-540-153-2	Sequence 2, Appli
21	46.5	24.2	1529	4	US-09-312-283C-396	Sequence 396, App
22	45.5	23.7	196	3	US-09-129-030-28	Sequence 28, Appl
23	45.5	23.7	694	4	US-09-252-991A-22481	Sequence 22481, A
24	45	23.4	477	4	US-09-252-991A-16778	Sequence 16778, A
25	45	23.4	2860	2	US-08-826-267-2	Sequence 2, Appli
26	44.5	23.2	270	4	US-09-399-913-59	Sequence 59, Appl
27	44.5	23.2	401	1	US-08-368-803-7	Sequence 7, Appli
28	44.5	23.2	422	2	US-08-663-566A-5	Sequence 5, Appli
29	44.5	23.2	422	2	US-08-023-610-5	Sequence 5, Appli
30	44.5	23.2	422	2	US-08-288-065A-5	Sequence 5, Appli
31	44.5	23.2	422	2	US-08-362-240A-5	Sequence 5, Appli
32	44.5	23.2	422	3	US-08-804-372A-3	Sequence 3, Appli
33	44.5	23.2	422	5	PCT-US95-10245-5	Sequence 5, Appli
34	44.5	23.2	442	4	US-09-252-991A-26462	Sequence 26462, A
35	44	22.9	113	3	US-08-836-075A-78	Sequence 78, Appl
36	44	22.9	179	4	US-09-252-991A-18885	Sequence 18885, A
37	44	22.9	222	6	5223425-6	Patent No. 5223425
38	44	22.9	228	3	US-08-944-483-44	Sequence 44, Appl
39	44	22.9	238	6	5223425-5	Patent No. 5223425
40	44	22.9	250	6	5223425-4	Patent No. 5223425
41	44	22.9	253	6	5223425-8	Patent No. 5223425
42	44	22.9	260	4	US-09-252-991A-17498	Sequence 17498, A
43	44	22.9	425	1	US-08-190-802A-48	Sequence 48, Appl
44	44	22.9	425	3	US-08-477-346-48	Sequence 48, Appl
45	44	22.9	425	4	US-08-473-089-48	Sequence 48, Appl

#### ALIGNMENTS

##### RESULT 1

US-08-928-862-4

; Sequence 4, Application US/08928862

; Patent No. 6309877

; GENERAL INFORMATION:

; APPLICANT: Chau, Raymond M. W.

; TITLE OF INVENTION: Isolation and Use of Motoneuronotrophic Factors

; FILE REFERENCE: 12592-2

; CURRENT APPLICATION NUMBER: US/08/928,862

; CURRENT FILING DATE: 1997-09-12

; NUMBER OF SEQ ID NOS: 5

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 4

; LENGTH: 33

; TYPE: PRT

; ORGANISM: Homo sapiens

US-08-928-862-4

Qy 1 LGTFWGD<sup>1</sup>TLNCWMLS<sup>1</sup>SAFSRYARCLAE<sup>1</sup>GH<sup>1</sup>DGPTQ 33  
 |||  
 Db 1 LGTFWGD<sup>1</sup>TLNCWMLS<sup>1</sup>SAFSRYARCLAE<sup>1</sup>GH<sup>1</sup>DGPTQ 33

Query Match 27.1%; Score 52; DB 4; Length 385;  
Best Local Similarity 52.4%; Pred. No. 18;  
Matches 11; Conservative 4; Mismatches 6; Indels 0; Gaps 0;

Qy 1 LGTFWGD~~T~~LNCWMLSAFSRYA 21  
:| | ||: :||:| |||  
Db 215 IGAFDGD~~T~~VKKFMLAARHRYA 235

```

US-08-914-375C-60
; Sequence 60, Application US/08914375C
; Patent No. 6377893
;   GENERAL INFORMATION:
;       APPLICANT: Steven A. Benner
;           Applications of Protein Structure Predictions
;       NUMBER OF SEQUENCES: 74
;       CORRESPONDENCE ADDRESS:
;           ADDRESSEE: Steven A. Benner
;           STREET: 1501 NW 68th Terrace
;           CITY: Gainesville
;           STATE: FL
;           COUNTRY: United States
;           ZIP: 32605-4147
;       COMPUTER READABLE FORM:

```

```

;          MEDIUM TYPE: 3.5 inch diskette
;          COMPUTER: Apple MacIntosh
;          OPERATING SYSTEM: MacIntosh 7.0
;          SOFTWARE: Microsoft Word
;    CURRENT APPLICATION DATA:
;          APPLICATION NUMBER: US/08/914,375C
;          FILING DATE: 19-Aug-1997
;          CLASSIFICATION: 702/20
;    TELECOMMUNICATION INFORMATION:
;          TELEPHONE: 352 392 7773
;          TELEFAX: 352 331 0462
;    INFORMATION FOR SEQ ID NO: 60:
;          SEQUENCE CHARACTERISTICS:
;            LENGTH: 466
;            TYPE: amino acid
;            TOPOLOGY: linear
;          MOLECULE TYPE: amino acid
;          ORIGINAL SOURCE:
;            ORGANISM: Escherichia coli
;          FEATURE:
;            OTHER INFORMATION: ascb_ecoli 6-phospho-strand-glucosidase (E.C.
3.2.1.86)
;          SEQUENCE DESCRIPTION: SEQ ID NO: 60:
US-08-914-375C-60

```

```

Query Match          26.6%;  Score 51;  DB 4;  Length 466;
Best Local Similarity 43.3%;  Pred. No. 31;
Matches    13;  Conservative    3;  Mismatches    14;  Indels      0;  Gaps      0;

```

```

Qy          1 LGTFWGDTLNCWMLSAFSRYARCLAEGHDG 30
             | | :| | | :| | | | | | | | |
Db          138 LVTEYGSWRNRKLVEFFSRYARTCFEAFDG 167

```

#### RESULT 4

```

US-09-328-352-6124
; Sequence 6124, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
ACINETOBACTER
; TITLE OF INVENTION: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC99-03PA
; CURRENT APPLICATION NUMBER: US/09/328,352
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 6124
; LENGTH: 341
; TYPE: PRT
; ORGANISM: Acinetobacter baumannii
US-09-328-352-6124

```

```

Query Match          25.5%;  Score 49;  DB 4;  Length 341;
Best Local Similarity 28.6%;  Pred. No. 41;
Matches     8;  Conservative    6;  Mismatches    14;  Indels      0;  Gaps      0;

```



Qy 1 LGTFWGDTLNCWMLSAFSRYARCLAEGH 28  
:| : | ||| :| : : |:  
Db 216 VGGLEFFDDLNCWDFETCFKYIQA VNGY 243

RESULT 5

US-08-836-075A-80

; Sequence 80, Application US/08836075A

; Patent No. 6180768

; GENERAL INFORMATION:

; APPLICANT: MAERTENS, GEERT

; APPLICANT: STUYVER, LIEVEN

; TITLE OF INVENTION: NEW SEQUENCES OF HEPATITIS C VIRUS GENOTYPES

; TITLE OF INVENTION: AND THEIR USE AS PROPHYLACTIC, THERAPEUTIC AND

DIAGNOSTIC

; TITLE OF INVENTION: AGENTS

; NUMBER OF SEQUENCES: 207

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: ARNOLD, WHITE & DURKEE

; STREET: P.O. BOX 4433

; CITY: HOUSTON

; STATE: TEXAS

; COUNTRY: USA

; ZIP: 77210-4433

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Microsoft Word 6.0 / ASCII text output

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/836,075A

; FILING DATE: 21 Apr 1997

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: PCT/EP95/04155

; FILING DATE: 23 Oct 1995

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: EP 94870166.9

; FILING DATE: 21 Oct 1994

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: EP 95870076.7

; FILING DATE: 28 Jun 1995

; ATTORNEY/AGENT INFORMATION:

; NAME: KAMMERER, PATRICIA A.

; REGISTRATION NUMBER: 29,775

; REFERENCE/DOCKET NUMBER: INNS:004

; INFORMATION FOR SEQ ID NO: 80:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 113 amino acids

; TYPE: amino acid

; TOPOLOGY: linear

; MOLECULE TYPE: peptide

US-08-836-075A-80

Query Match 25.0%; Score 48; DB 3; Length 113;

Best Local Similarity 36.7%; Pred. No. 16;

Matches 11; Conservative 4; Mismatches 11; Indels 4; Gaps 1;

Qy 3 TFWGDTLNCWMLSAFSRYARCLAEGHDGPT 32  
| |:|: |:|: | | | | | |  
Db 62 TSMGNTITCYV----KMAAACRAAGIDAPT 87

RESULT 6

US-08-644-271-30

; Sequence 30, Application US/08644271

; Patent No. 5814478

; GENERAL INFORMATION:

; APPLICANT: Valenzuela, et al.

; TITLE OF INVENTION: NOVEL TYROSINE KINASE RECEPTORS

; TITLE OF INVENTION: AND LIGANDS

; NUMBER OF SEQUENCES: 32

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Regeneron Pharmaceuticals, Inc.

; STREET: 777 Old Saw Mill Road

; CITY: Tarrytown

; STATE: NY

; COUNTRY: USA

; ZIP: 10591

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette

; COMPUTER: IBM Compatible

; OPERATING SYSTEM: DOS

; SOFTWARE: FastSEQ Version 2.0

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/644,271

; FILING DATE: 10-MAY-1996

; CLASSIFICATION: 435

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: USSN 60/008,657

; FILING DATE: 15-DEC-1995

; ATTORNEY/AGENT INFORMATION:

; NAME: Cobert, Robert J

; REGISTRATION NUMBER: 36,108

; REFERENCE/DOCKET NUMBER: REG 195A

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 914-345-7400

; TELEFAX: 914-345-7721

; TELEX:

; INFORMATION FOR SEQ ID NO: 30:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 1940 amino acids

; TYPE: amino acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: protein

; FEATURE:

; NAME/KEY: Rat Agrin

; LOCATION: 1...1940

; OTHER INFORMATION:

US-08-644-271-30

Query Match 25.0%; Score 48; DB 2; Length 1940;

Best Local Similarity 39.3%; Pred. No. 4.1e+02;

Matches 11; Conservative 1; Mismatches 14; Indels 2; Gaps 1;

Qy            6 GDTLN--CWMLSAFSRYARCLAEHGDGP 31  
              | | | | | | | : | | |  
 Db           328 GHTYNDCWRQQAECRQQRAIPPKHQGP 355

RESULT 7

US-09-077-955-34  
 ; Sequence 34, Application US/09077955A  
 ; Patent No. 6413740  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Valenzuela et al., David M.  
 ; TITLE OF INVENTION: NOVEL TYROSINE KINASE RECEPTORS AND LIGANDS  
 ; FILE REFERENCE: REG195-B-PCT-US  
 ; CURRENT APPLICATION NUMBER: US/09/077,955A  
 ; CURRENT FILING DATE: 1998-09-10  
 ; EARLIER APPLICATION NUMBER: PCT/US96/20696  
 ; EARLIER FILING DATE: 1996-12-13  
 ; EARLIER APPLICATION NUMBER: 08/644,271  
 ; EARLIER FILING DATE: 1996-05-10  
 ; EARLIER APPLICATION NUMBER: 60/008,657  
 ; EARLIER FILING DATE: 1995-12-15  
 ; NUMBER OF SEQ ID NOS: 36  
 ; SOFTWARE: PatentIn Ver. 2.0  
 ; SEQ ID NO 34  
 ; LENGTH: 1940  
 ; TYPE: PRT  
 ; ORGANISM: Rattus sp.  
 US-09-077-955-34

Query Match            25.0%; Score 48; DB 4; Length 1940;  
 Best Local Similarity   39.3%; Pred. No. 4.1e+02;  
 Matches   11; Conservative   1; Mismatches   14; Indels   2; Gaps   1;

Qy            6 GDTLN--CWMLSAFSRYARCLAEHGDGP 31  
              | | | | | | | : | | |  
 Db           328 GHTYNDCWRQQAECRQQRAIPPKHQGP 355

RESULT 8

US-09-252-991A-17047  
 ; Sequence 17047, Application US/09252991A  
 ; Patent No. 6551795  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Marc J. Rubenfield et al.  
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO  
 PSEUDOMONAS  
 ; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
 ; FILE REFERENCE: 107196.136  
 ; CURRENT APPLICATION NUMBER: US/09/252,991A  
 ; CURRENT FILING DATE: 1999-02-18  
 ; PRIOR APPLICATION NUMBER: US 60/074,788  
 ; PRIOR FILING DATE: 1998-02-18  
 ; PRIOR APPLICATION NUMBER: US 60/094,190  
 ; PRIOR FILING DATE: 1998-07-27  
 ; NUMBER OF SEQ ID NOS: 33142  
 ; SEQ ID NO 17047

; LENGTH: 477  
; TYPE: PRT  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-17047

Query Match 24.7%; Score 47.5; DB 4; Length 477;  
Best Local Similarity 45.5%; Pred. No. 97;  
Matches 10; Conservative 2; Mismatches 7; Indels 3; Gaps 1;

Qy 11 CW---MLSAFSRYARCLAEGHD 29  
|| : |: | |||||  
Db 202 CWEEHSRNMFAHMERTLAEGHD 223

RESULT 9

US-09-612-204B-24  
; Sequence 24, Application US/09612204B  
; Patent No. 6461811  
; GENERAL INFORMATION:  
; APPLICANT: Patience, Clive  
; TITLE OF INVENTION: Swine Gamma Herpesvirus DNA and Methods of Use  
; FILE REFERENCE: 61750-299  
; CURRENT APPLICATION NUMBER: US/09/612,204B  
; CURRENT FILING DATE: 2001-08-13  
; PRIOR APPLICATION NUMBER: U.S. 60/142,736  
; PRIOR FILING DATE: 1999-07-08  
; PRIOR APPLICATION NUMBER: U.S. 60/168,532  
; PRIOR FILING DATE: 1999-12-02  
; NUMBER OF SEQ ID NOS: 36  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 24  
; LENGTH: 865  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence:Deduced amino  
; OTHER INFORMATION: acid sequence of porcine gamma herpesvirus gpB  
; OTHER INFORMATION: gene  
US-09-612-204B-24

Query Match 24.7%; Score 47.5; DB 4; Length 865;  
Best Local Similarity 39.1%; Pred. No. 1.9e+02;  
Matches 9; Conservative 4; Mismatches 5; Indels 5; Gaps 1;

Qy 2 GTFWGD-----TLNCWMLSAFSR 19  
| ||| |:|| :: |:|  
Db 228 GFWGGSYRRRTTVNCELMDFAR 250

RESULT 10

US-09-252-991A-30553  
; Sequence 30553, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO  
PSEUDOMONAS

; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; CURRENT FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 30553  
; LENGTH: 230  
; TYPE: PRT  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-30553

Query Match 24.5%; Score 47; DB 4; Length 230;  
Best Local Similarity 58.3%; Pred. No. 50;  
Matches 7; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Qy 11 CWMLSAFSRYAR 22  
||:|| || :|  
Db 130 CWVLSCFSSFSR 141

RESULT 11

US-09-252-991A-22859  
; Sequence 22859, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO  
PSEUDOMONAS  
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; CURRENT FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 22859  
; LENGTH: 736  
; TYPE: PRT  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-22859

Query Match 24.5%; Score 47; DB 4; Length 736;  
Best Local Similarity 32.3%; Pred. No. 1.9e+02;  
Matches 10; Conservative 3; Mismatches 18; Indels 0; Gaps 0;

Qy 1 LGTFWGDTLNCWMLSAFSRYARCLAEHGDGP 31  
| |: | |:| | | |: |  
Db 687 LETYLTDTNTQAWVLQADGSYQRLSPTGNQNP 717

RESULT 12

US-09-252-991A-18159  
 ; Sequence 18159, Application US/09252991A  
 ; Patent No. 6551795  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Marc J. Rubenfield et al.  
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO  
 PSEUDOMONAS  
 ; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
 ; FILE REFERENCE: 107196.136  
 ; CURRENT APPLICATION NUMBER: US/09/252,991A  
 ; CURRENT FILING DATE: 1999-02-18  
 ; PRIOR APPLICATION NUMBER: US 60/074,788  
 ; PRIOR FILING DATE: 1998-02-18  
 ; PRIOR APPLICATION NUMBER: US 60/094,190  
 ; PRIOR FILING DATE: 1998-07-27  
 ; NUMBER OF SEQ ID NOS: 33142  
 ; SEQ ID NO 18159  
 ; LENGTH: 1778  
 ; TYPE: PRT  
 ; ORGANISM: Pseudomonas aeruginosa  
 US-09-252-991A-18159

Query Match 24.5%; Score 47; DB 4; Length 1778;  
 Best Local Similarity 54.5%; Pred. No. 5.1e+02;  
 Matches 12; Conservative 1; Mismatches 3; Indels 6; Gaps 1;

Qy 12 WMLSA-----FSRYARCLAEG 27  
 | ||| ||:||| | ||  
 Db 1120 WELSAERPFTFSKYARQLQEG 1141

RESULT 13  
 US-09-191-647-14  
 ; Sequence 14, Application US/09191647  
 ; Patent No. 6046015  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Goodman, Corey  
 ; APPLICANT: Kid, Thomas  
 ; APPLICANT: Brose, Katja  
 ; APPLICANT: Tessier-Lavigne, Marc  
 ; TITLE OF INVENTION: Modulating Robo: Ligand Interactions  
 ; FILE REFERENCE: B98-031-3  
 ; CURRENT APPLICATION NUMBER: US/09/191,647  
 ; CURRENT FILING DATE: 1998-11-13  
 ; EARLIER APPLICATION NUMBER: 60/065,544  
 ; EARLIER FILING DATE: 1997-11-14  
 ; EARLIER APPLICATION NUMBER: 60/081,057  
 ; EARLIER FILING DATE: 1998-04-07  
 ; NUMBER OF SEQ ID NOS: 14  
 ; SOFTWARE: PatentIn Ver. 2.0  
 ; SEQ ID NO 14  
 ; LENGTH: 243  
 ; TYPE: PRT  
 ; ORGANISM: mouse  
 US-09-191-647-14

Query Match 24.2%; Score 46.5; DB 3; Length 243; .

Best Local Similarity 50.0%; Pred. No. 62;  
Matches 10; Conservative 3; Mismatches 6; Indels 1; Gaps 1;

Qy 11 CWMLSAFSRYARCLAEGHDG 30  
| :||| :|| ||| |  
Db 99 CLPINAFSYSCKCL-EGHGG 117

RESULT 14

US-09-540-245A-14

; Sequence 14, Application US/09540245A  
; Patent No. 6270984  
; GENERAL INFORMATION:  
; APPLICANT: Goodman, Corey  
; APPLICANT: Kid, Thomas  
; APPLICANT: Brose, Katja  
; APPLICANT: Tessier-Lavigne, Marc  
; TITLE OF INVENTION: Modulating Robo: Ligand Interactions  
; FILE REFERENCE: B98-031-3  
; CURRENT APPLICATION NUMBER: US/09/540,245A  
; CURRENT FILING DATE: 2000-03-31  
; PRIOR APPLICATION NUMBER: 60/065,544  
; PRIOR FILING DATE: 1997-11-14  
; PRIOR APPLICATION NUMBER: 60/081,057  
; PRIOR FILING DATE: 1998-04-07  
; NUMBER OF SEQ ID NOS: 20  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 14  
; LENGTH: 243  
; TYPE: PRT  
; ORGANISM: mouse  
US-09-540-245A-14

Query Match 24.2%; Score 46.5; DB 3; Length 243;  
Best Local Similarity 50.0%; Pred. No. 62;  
Matches 10; Conservative 3; Mismatches 6; Indels 1; Gaps 1;

Qy 11 CWMLSAFSRYARCLAEGHDG 30  
| :||| :|| ||| |  
Db 99 CLPINAFSYSCKCL-EGHGG 117

RESULT 15

US-09-540-153-14

; Sequence 14, Application US/09540153  
; Patent No. 6270995  
; GENERAL INFORMATION:  
; APPLICANT: Goodman, Corey  
; APPLICANT: Kid, Thomas  
; APPLICANT: Brose, Katja  
; APPLICANT: Tessier-Lavigne, Marc  
; TITLE OF INVENTION: Modulating Robo: Ligand Interactions  
; FILE REFERENCE: B98-031-3  
; CURRENT APPLICATION NUMBER: US/09/540,153  
; CURRENT FILING DATE: 2000-03-31  
; PRIOR APPLICATION NUMBER: 09/191,647  
; PRIOR FILING DATE: 1998-11-13

; PRIOR APPLICATION NUMBER: 60/081,057  
; PRIOR FILING DATE: 1998-04-07  
; NUMBER OF SEQ ID NOS: 14  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 14  
; LENGTH: 243  
; TYPE: PRT  
; ORGANISM: mouse  
US-09-540-153-14

Query Match 24.2%; Score 46.5; DB 3; Length 243;  
Best Local Similarity 50.0%; Pred. No. 62;  
Matches 10; Conservative 3; Mismatches 6; Indels 1; Gaps 1;

Qy 11 CWMLSAFSRYARCLAEGHDG 30  
| :||| :|| ||| |  
Db 99 CLPINAFSYSCKCL-EGHGG 117

Search completed: January 30, 2004, 11:27:44  
Job time : 2.86187 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 30, 2004, 11:21:12 ; Search time 1.79767 Seconds  
(without alignments)  
1765.382 Million cell updates/sec

Title: US-09-989-481-4  
Perfect score: 192  
Sequence: 1 LGTFWGDTLNCWMLSAFSRYARCLAEHGDGPTQ 33

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0  
Maximum DB seq length: 20000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR\_76:\*  
1: pir1:\*  
2: pir2:\*  
3: pir3:\*  
4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query		DB	ID	Description
		Match	Length			
1	59	30.7	910	2	T38539	probable importin
2	56.5	29.4	223	2	D70760	hypothetical prote
3	55	28.6	694	2	T52574	cyclic nucleotide
4	54	28.1	287	2	T39197	yeast atp12 protei
5	54	28.1	316	2	C87318	hypothetical prote
6	52	27.1	255	2	G87668	conserved hypothet
7	52	27.1	533	2	T01864	hypothetical prote
8	52	27.1	544	2	S41090	triacylglycerol li
9	52	27.1	563	1	ACGUGC	triacylglycerol li
10	51.5	26.8	1150	2	S58775	mypl protein - smu
11	51	26.6	90	2	S24248	Ig heavy chain V r
12	51	26.6	101	2	S24257	Ig heavy chain V r
13	51	26.6	105	2	S24249	Ig heavy chain V r

14	51	26.6	109	2	S24254	Ig heavy chain V r
15	51	26.6	109	2	S24253	Ig heavy chain V r
16	51	26.6	110	2	S24250	Ig heavy chain V r
17	51	26.6	113	2	S24247	Ig heavy chain V r
18	51	26.6	373	2	G85355	nodulin-like prote
19	51	26.6	418	2	S11678	cyclin A - African
20	51	26.6	472	2	A75464	probable zinc meta
21	51	26.6	474	2	D91075	6-phospho-beta-glu
22	51	26.6	474	2	C85920	6-phospho-beta-glu
23	51	26.6	474	2	H65051	6-phospho-beta-glu
24	51	26.6	1207	1	EGHU	epidermal growth f
25	50.5	26.3	537	2	T05816	hypothetical prote
26	50	26.0	726	2	T51519	cyclic nucleotide-
27	49.5	25.8	355	2	T50479	G protein alpha ch
28	49.5	25.8	408	2	T38386	hypothetical wd-40
29	49	25.5	117	2	T46430	hypothetical prote
30	49	25.5	414	2	H86736	rod-shape determin
31	49	25.5	709	2	D82825	conserved hypothet
32	48.5	25.3	293	2	D95919	hypothetical prote
33	48.5	25.3	506	2	T35261	probable metallope
34	48	25.0	135	2	H71051	hypothetical prote
35	48	25.0	229	2	D65209	peptidase E - Esch
36	48	25.0	229	2	C91246	peptidase E [impor
37	48	25.0	229	2	A86094	peptidase E [impor
38	48	25.0	353	2	AC3320	integral membrane
39	48	25.0	1959	1	AGRT	agrin - rat
40	47.5	24.7	339	2	T09217	protein sam2B - sp
41	47.5	24.7	467	2	G83592	adenosylmethionine
42	47.5	24.7	477	2	T05202	pectinesterase hom
43	47	24.5	102	2	S24260	Ig heavy chain V r
44	47	24.5	115	2	T31781	hypothetical prote
45	47	24.5	241	2	C82852	competence protein

#### ALIGNMENTS

##### RESULT 1

T38539

probable importin beta-2 subunit (transportin) - fission yeast  
(Schizosaccharomyces pombe)

C;Species: Schizosaccharomyces pombe

C;Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 03-Dec-1999

C;Accession: T38539

R;Oliver, K.; Harris, D.; Barrell, B.G.; Rajandream, M.A.; Wood, V.  
submitted to the EMBL Data Library, September 1997

A;Reference number: Z21748

A;Accession: T38539

A;Status: preliminary; translated from GB/EMBL/DDBJ

A;Molecule type: DNA

A;Residues: 1-910 <OLI>

A;Cross-references: EMBL:Z99165; PIDN:CAB16272.1; GSPDB:GN00066;  
SPDB:SPAC2F3.06c

A;Experimental source: strain 972h-; cosmid c2F3

C;Genetics:

A;Gene: SPDB:SPAC2F3.06c

A;Map position: 1

A;Introns: 36/3

Query Match 30.7%; Score 59; DB 2; Length 910;  
Best Local Similarity 40.9%; Pred. No. 3.8;  
Matches 9; Conservative 4; Mismatches 9; Indels 0; Gaps 0;

Qy 8 TLNCWMLSAFSRYARCLAEAGHD 29  
|: || | :|::| || |  
Db 473 TITCWTLGRYSKWASCLESEED 494

RESULT 2

D70760

hypothetical protein Rv2014 - Mycobacterium tuberculosis (strain H37RV)

C;Species: Mycobacterium tuberculosis

C;Date: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 22-Oct-1999

C;Accession: D70760

R;Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.;  
Gordon, S.V.; Eiglmeier, K.; Gas, S.; Barry III, C.E.; Tekaiia, F.; Badcock, K.;  
Basham, D.; Brown, D.; Chillingworth, T.; Connor, R.; Davies, R.; Devlin, K.;  
Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Hornsby, T.; Jagels, K.;  
Krogh, A.; McLean, J.; Moule, S.; Murphy, L.; Oliver, S.; Osborne, J.; Quail,  
M.A.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.;  
Squares, S.

Nature 393, 537-544, 1998

A;Authors: Sqaers, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.

A;Title: Deciphering the biology of Mycobacterium tuberculosis from the complete  
genome sequence.

A;Reference number: A70500; MUID:98295987; PMID:9634230

A;Accession: D70760

A;Status: preliminary; nucleic acid sequence not shown; translation not shown

A;Molecule type: DNA

A;Residues: 1-223 <COL>

A;Cross-references: GB:Z74025; GB:AL123456; NID:g3261586; PIDN:CAA98415.1;

PID:e1299911; PID:g3261592

A;Experimental source: strain H37Rv

C;Genetics:

A;Gene: Rv2014

Query Match 29.4%; Score 56.5; DB 2; Length 223;  
Best Local Similarity 46.7%; Pred. No. 2;  
Matches 14; Conservative 2; Mismatches 9; Indels 5; Gaps 2;

Qy 4 FWGDT--LNCWMLSAFSRYARCLAEAGHDGP 31  
| ||: | | | || | : | || |  
Db 157 FAGDSRRANLW---AADRYNRAIARGHDHP 183

RESULT 3

T52574

cyclic nucleotide and calmodulin-regulated ion channel [imported] - Arabidopsis  
thaliana

C;Species: Arabidopsis thaliana (mouse-ear cress)

C;Date: 24-Oct-2000 #sequence\_revision 24-Oct-2000 #text\_change 24-Oct-2000

C;Accession: T52574

R;Kohler, C.; Merkle, T.; Neuhaus, G.

Plant J. 18, 97-104, 1999

A;Title: Characterisation of a novel gene family of putative cyclic nucleotide- and calmodulin-regulated ion channels in *Arabidopsis thaliana*.  
A;Reference number: Z26120  
A;Accession: T52574  
A;Status: preliminary; translated from GB/EMBL/DDBJ  
A;Molecule type: mRNA  
A;Residues: 1-694 <KOH>  
A;Cross-references: EMBL:Y17912; PIDN:CAB40129.1  
A;Experimental source: cultivar Columbia  
C;Genetics:  
A;Gene: cngc4

Query Match 28.6%; Score 55; DB 2; Length 694;  
Best Local Similarity 35.9%; Pred. No. 10;  
Matches 14; Conservative 2; Mismatches 9; Indels 14; Gaps 2;

Qy 2 GT-FWGDTLN-----CWMLSAFSRYARCLAE 26  
|| :|| || || | | :|| |  
Db 255 GTVWWGIALNMIAYFVAHAAGACWYLLGVQRSACKLKE 293

#### RESULT 4

T39197

yeast atp12 protein precursor homolog - fission yeast (*Schizosaccharomyces pombe*)

C;Species: *Schizosaccharomyces pombe*

C;Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 03-Dec-1999

C;Accession: T39197

R;Wedler, H.; Duesterhoeft, A.; Lyne, M.H.; Rajandream, M.A.; Barrell, B.G. submitted to the EMBL Data Library, October 1999

A;Reference number: Z21834

A;Accession: T39197

A;Status: preliminary; translated from GB/EMBL/DDBJ

A;Molecule type: DNA

A;Residues: 1-287 <WED>

A;Cross-references: EMBL:AL121764; PIDN:CAB57430.1; GSPDB:GN00066; SPDB:SPAC9.12c

A;Experimental source: strain 972h-; cosmid c9

C;Genetics:

A;Gene: SPDB:SPAC9.12c

A;Map position: 1

A;Introns: 257/1

Query Match 28.1%; Score 54; DB 2; Length 287;  
Best Local Similarity 47.4%; Pred. No. 5.8;  
Matches 9; Conservative 2; Mismatches 8; Indels 0; Gaps 0;

Qy 5 WGDTLNCWMLSAFSRYARC 23  
| :|| | |:|| | |  
Db 198 WLSSLNSWQLAAFSRVSC 216

#### RESULT 5

C87318

hypothetical protein CC0557 [imported] - *Caulobacter crescentus*

C;Species: *Caulobacter crescentus*

C;Date: 20-Apr-2001 #sequence\_revision 20-Apr-2001 #text\_change 20-Apr-2001

C;Accession: C87318  
 R;Niernan, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, J.F.; Alley, M.; Ohta, N.; Maddock, J.R.; Potocka, I.; Nelson, W.C.; Newton, A.; Stephens, C.; Phadke, N.D.; Ely, B.; Laub, M.T.; DeBoy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Kolonay, J.F.; Smit, J.; Craven, M.; Khouri, H.; Shetty, J.; Berry, K.; Utterback, T.; Tran, K.; Wolf, A.; Vamathevan, J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.M.  
 Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001  
 A;Title: Complete Genome Sequence of *Caulobacter crescentus*.  
 A;Reference number: A87249; MUID:21173698; PMID:11259647  
 A;Accession: C87318  
 A;Status: preliminary  
 A;Molecule type: DNA  
 A;Residues: 1-316 <STO>  
 A;Cross-references: GB:AE005673; NID:g13421749; PIDN:AAK22543.1; GSPDB:GN00148  
 C;Genetics:  
 A;Gene: CC0557

Query Match 28.1%; Score 54; DB 2; Length 316;  
 Best Local Similarity 56.2%; Pred. No. 6.4;  
 Matches 9; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

Qy 6 GDTLNCWMLSAFSRYA 21  
 || |:|| | | ||:  
 Db 301 GDILSCWKLGAVERYS 316

RESULT 6  
 G87668  
 conserved hypothetical protein CC3385 [imported] - *Caulobacter crescentus*  
 C;Species: *Caulobacter crescentus*  
 C;Date: 20-Apr-2001 #sequence\_revision 20-Apr-2001 #text\_change 20-Apr-2001  
 C;Accession: G87668  
 R;Niernan, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, J.F.; Alley, M.; Ohta, N.; Maddock, J.R.; Potocka, I.; Nelson, W.C.; Newton, A.; Stephens, C.; Phadke, N.D.; Ely, B.; Laub, M.T.; DeBoy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Kolonay, J.F.; Smit, J.; Craven, M.; Khouri, H.; Shetty, J.; Berry, K.; Utterback, T.; Tran, K.; Wolf, A.; Vamathevan, J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.M.  
 Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001  
 A;Title: Complete Genome Sequence of *Caulobacter crescentus*.  
 A;Reference number: A87249; MUID:21173698; PMID:11259647  
 A;Accession: G87668  
 A;Status: preliminary  
 A;Molecule type: DNA  
 A;Residues: 1-255 <STO>  
 A;Cross-references: GB:AE005673; NID:g13425093; PIDN:AAK25347.1; GSPDB:GN00148  
 C;Genetics:  
 A;Gene: CC3385

Query Match 27.1%; Score 52; DB 2; Length 255;  
 Best Local Similarity 41.4%; Pred. No. 9.8;  
 Matches 12; Conservative 2; Mismatches 7; Indels 8; Gaps 1;

Qy 4 FWGDTLNCWMLSAFSRYARCLAEHGDGPT 32

Db                    |||:|:            ||        |||| ||  
144 FWGETI-----SRTLNQAAEGHADPT 164

RESULT 7

T01864

hypothetical protein T7M24.1 - Arabidopsis thaliana

C;Species: Arabidopsis thaliana (mouse-ear cress)

C;Date: 26-Feb-1999 #sequence\_revision 26-Feb-1999 #text\_change 24-Mar-1999

C;Accession: T01864

R;Harmon, G.; Langston, Y.; Stoneking, T.; Drone, K.; Ames, M.  
submitted to the EMBL Data Library, July 1998

A;Description: The sequence of Arabidopsis thaliana T7M24.

A;Reference number: Z14448

A;Accession: T01864

A;Status: translated from GB/EMBL/DDBJ

A;Molecule type: DNA

A;Residues: 1-533 <HAR>

A;Cross-references: EMBL:AF077408; NID:g3319359; PID:g3319364

A;Experimental source: cultivar Columbia

C;Genetics:

A;Map position: 4

A;Introns: 92/3; 105/2; 152/3; 268/1; 381/3

A;Note: T7M24.1

Query Match                    27.1%;    Score 52;    DB 2;    Length 533;  
Best Local Similarity        34.5%;    Pred. No. 21;  
Matches    10;    Conservative        6;    Mismatches    13;    Indels        0;    Gaps        0;

Qy                    1 LGTFWGDTLNCWMLSAFSRYARCLAEGHD 29

|| : ::| ||        :|| | | |

Db                    145 LGQIYKESVNYWMSHRTLKFARHLVRGRD 173

RESULT 8

S41090

triacylglycerol lipase (EC 3.1.1.3) I precursor - yeast (Geotrichum candidum)  
(strain ATCC 34614)

C;Species: Geotrichum candidum

A;Variety: ATCC 34614

C;Date: 19-Mar-1997 #sequence\_revision 05-Feb-1999 #text\_change 18-Jun-1999

C;Accession: S41090

R;Bertolini, M.C.; Laramée, L.; Thomas, D.Y.; Cygler, M.; Schrag, J.D.; Vernet,  
T.

Eur. J. Biochem. 219, 119-125, 1994

A;Title: Polymorphism in the lipase genes of Geotrichum candidum strains.

A;Reference number: S41090; MUID:94139683; PMID:8306978

A;Accession: S41090

A;Status: nucleic acid sequence not shown; not compared with conceptual  
translation

A;Molecule type: DNA

A;Residues: 1-544 <BER>

A;Cross-references: GB:U02622; NID:g409275; PIDN:AAA03435.1; PID:g409276

A;Experimental source: ATCC 34614

A;Note: only the translation of the mature protein is shown

C;Genetics:

A;Gene: lipI

C;Function:

A;Description: hydrolyzes triacylglycerols into fatty acids and glycerol

C;Superfamily: cholinesterase; cholinesterase homology

C;Keywords: carboxylic ester hydrolase; glycoprotein; lipid hydrolysis; pyroglutamic acid

F;24-541/Domain: cholinesterase homology <CHE>

F;215-219/Region: interfacial lipid recognition (GX SXG) motif

F;1/Modified site: pyrrolidone carboxylic acid (Gln) #status predicted

F;61-105,276-288/Disulfide bonds: #status predicted

F;217/Active site: Ser #status predicted

F;283,364/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 27.1%; Score 52; DB 2; Length 544;

Best Local Similarity 40.0%; Pred. No. 21;

Matches 14; Conservative 1; Mismatches 14; Indels 6; Gaps 1;

Qy 1 LGTFWGD TL-----NCWMLSAFSRYARCLAEGHD 29

|||| | | | | | ||: || | ||

Db 459 LGTFHGS D L L F Q Y Y A G P W S S S A Y R R Y F I S F A N H H D 493

RESULT 9

ACGUGC

triacylglycerol lipase (EC 3.1.1.3) I precursor - yeast (*Geotrichum candidum*)

N;Alternate names: lipase

C;Species: *Geotrichum candidum*

C;Date: 31-Mar-1990 #sequence\_revision 31-Mar-1990 #text\_change 31-Mar-2000

C;Accession: PN0492; JQ0022

R;Nagao, T.; Shimada, Y.; Sugihara, A.; Tominaga, Y.

J. Biochem. 113, 776-780, 1993

A;Title: Cloning and sequencing of two chromosomal lipase genes from *Geotrichum candidum*.

A;Reference number: PN0492; MUID:93380907; PMID:8370674

A;Accession: PN0492

A;Molecule type: DNA

A;Residues: 1-563 <NAG>

A;Note: the translation of residues 31-550 and the corresponding nucleotide sequence are not shown

R;Shimada, Y.; Sugihara, A.; Tominaga, Y.; Iizumi, T.; Tsunasawa, S.

J. Biochem. 106, 383-388, 1989

A;Title: cDNA molecular cloning of *Geotrichum candidum* lipase.

A;Reference number: JQ0022; MUID:90110016; PMID:2481674

A;Accession: JQ0022

A;Molecule type: mRNA

A;Residues: 1-563 <SHI>

A;Experimental source: strain ATCC 34614

A;Note: sequences of several small peptides were also determined

C;Comment: The extracellular lipase produced by *Geotrichum candidum* hydrolyzes all ester bonds in triglyceride and displays a high affinity for triolein.

C;Genetics:

A;Gene: lipI

C;Superfamily: cholinesterase; cholinesterase homology

C;Keywords: carboxylic ester hydrolase; glycoprotein; pyroglutamic acid

F;1-19/Domain: signal sequence #status predicted <SIG>

F;20-563/Product: triacylglycerol lipase #status experimental <MAT>

F;43-560/Domain: cholinesterase homology <CHE>

F;234-238/Region: interfacial lipid recognition (GX SXG) motif

F;20/Modified site: pyrrolidone carboxylic acid (Gln) (in mature form) #status experimental  
F;80-124,295-307/Disulfide bonds: #status predicted  
F;236/Active site: Ser #status predicted  
F;302,383/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 27.1%; Score 52; DB 1; Length 563;  
Best Local Similarity 40.0%; Pred. No. 22;  
Matches 14; Conservative 1; Mismatches 14; Indels 6; Gaps 1;

Qy 1 LGTFWGDTL-----NCWMLSAFSRYARCLAEGHD 29  
| | | | | | | | | | | | | | | | | | | | | |  
Db 478 LGTFHGSDLLEFQYYAGPWSSSAYRRYFISFANHHD 512

RESULT 10

S58775

mypl protein - smut fungus (*Ustilago maydis*)

C;Species: *Ustilago maydis* (corn smut)

C;Date: 12-Feb-1998 #sequence\_revision 20-Feb-1998 #text\_change 21-Jul-2000

C;Accession: S58775

R;Giasson, L.; Kronstad, J.W.

Genetics 141, 491-501, 1995

A;Title: Mutations in the mypl gene of *Ustilago maydis* attenuate mycelial growth and virulence.

A;Reference number: S58775; MUID:96109597; PMID:8647387

A;Accession: S58775

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-1150 <GIA>

A;Cross-references: EMBL:L33919; NID:g886415; PIDN:AAC37439.1; PID:g886416

C;Genetics:

A;Gene: mypl

Query Match 26.8%; Score 51.5; DB 2; Length 1150;  
Best Local Similarity 39.3%; Pred. No. 53;  
Matches 11; Conservative 7; Mismatches 7; Indels 3; Gaps 2;

Qy 1 LGTFWGDTLNCWMLSAFSRYARCLAEGH 28  
: | | | | : | | : | : | : | : | : |  
Db 714 IGTFW-LSRNAWILA--TRHGHLLSPGH 738

RESULT 11

S24248

Ig heavy chain V region (VH26) - human

C;Species: *Homo sapiens* (man)

C;Date: 19-Feb-1994 #sequence\_revision 10-Nov-1995 #text\_change 23-Jul-1999

C;Accession: S24248

R;Stewart, A.K.; Huang, C.; Stollar, B.D.; Schwartz, R.S.

submitted to the EMBL Data Library, June 1992

A;Description: A single VH gene predominates in the rearranged and expressed human B cell repertoires.

A;Reference number: S24247

A;Accession: S24248

A;Status: preliminary

A;Molecule type: DNA



A;Residues: 1-90 <STE>

A;Cross-references: EMBL:X67069; NID:g38395; PIDN:CAA47454.1; PID:g38396

C;Superfamily: immunoglobulin V region; immunoglobulin homology

C;Keywords: heterotetramer; immunoglobulin

Query Match 26.6%; Score 51; DB 2; Length 90;  
Best Local Similarity 52.2%; Pred. No. 4.7;  
Matches 12; Conservative 2; Mismatches 7; Indels 2; Gaps 1;

Qy 1 LGTFWG--DTLNCWMLSAFSRYA 21  
||| || :|| | : || ||  
Db 10 LGTAWGVPETLLCSLWFTFSSYA 32

RESULT 12

S24257

Ig heavy chain V region (VH26-DXP1-JH4) - human

C;Species: Homo sapiens (man)

C;Date: 19-Feb-1994 #sequence\_revision 10-Nov-1995 #text\_change 21-Jan-2000

C;Accession: S24257

R;Stewart, A.K.; Huang, C.; Stollar, B.D.; Schwartz, R.S.  
submitted to the EMBL Data Library, June 1992

A;Description: A single VH gene predominates in the rearranged and expressed  
human B cell repertoires.

A;Reference number: S24247

A;Accession: S24257

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-101 <STE>

A;Cross-references: EMBL:X67065; NID:g38387; PIDN:CAA47450.1; PID:g38388

C;Superfamily: immunoglobulin V region; immunoglobulin homology

C;Keywords: heterotetramer; immunoglobulin

F;11-93/Domain: immunoglobulin homology <IMM>

Query Match 26.6%; Score 51; DB 2; Length 101;  
Best Local Similarity 52.2%; Pred. No. 5.3;  
Matches 12; Conservative 2; Mismatches 7; Indels 2; Gaps 1;

Qy 1 LGTFWG--DTLNCWMLSAFSRYA 21  
||| || :|| | : || ||  
Db 6 LGTAWGVPETLLCSLWFTFSSYA 28

RESULT 13

S24249

Ig heavy chain V region (VH26-DN1-DXP1-JH4) - human

C;Species: Homo sapiens (man)

C;Date: 19-Feb-1994 #sequence\_revision 10-Nov-1995 #text\_change 30-May-1997

C;Accession: S24249

R;Stewart, A.K.; Huang, C.; Stollar, B.D.; Schwartz, R.S.  
submitted to the EMBL Data Library, June 1992

A;Description: A single VH gene predominates in the rearranged and expressed  
human B cell repertoires.

A;Reference number: S24247

A;Accession: S24249

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-105 <STE>  
A;Cross-references: EMBL:X67070  
C;Superfamily: immunoglobulin V region; immunoglobulin homology  
C;Keywords: heterotetramer; immunoglobulin  
F;10-92/Domain: immunoglobulin homology <IMM>

Query Match 26.6%; Score 51; DB 2; Length 105;  
Best Local Similarity 52.2%; Pred. No. 5.5;  
Matches 12; Conservative 2; Mismatches 7; Indels 2; Gaps 1;

Qy 1 LGTFWG--DTLNCWMLSAFSRYA 21  
||| || :|| | : || ||  
Db 5 LGTAWGVPETLLCSLWFTFSSYA 27

RESULT 14

S24254

Ig heavy chain V region (VH26-DXP2-JH4) - human

C;Species: Homo sapiens (man)

C;Date: 19-Feb-1994 #sequence\_revision 10-Nov-1995 #text\_change 21-Jan-2000

C;Accession: S24254

R;Stewart, A.K.; Huang, C.; Stollar, B.D.; Schwartz, R.S.

submitted to the EMBL Data Library, June 1992

A;Description: A single VH gene predominates in the rearranged and expressed human B cell repertoires.

A;Reference number: S24247

A;Accession: S24254

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-109 <STE>

A;Cross-references: EMBL:X67062

C;Superfamily: immunoglobulin V region; immunoglobulin homology

C;Keywords: heterotetramer; immunoglobulin

F;15-97/Domain: immunoglobulin homology <IMM>

Query Match 26.6%; Score 51; DB 2; Length 109;  
Best Local Similarity 52.2%; Pred. No. 5.7;  
Matches 12; Conservative 2; Mismatches 7; Indels 2; Gaps 1;

Qy 1 LGTFWG--DTLNCWMLSAFSRYA 21  
||| || :|| | : || ||  
Db 10 LGTAWGVPETLLCSLWFTFSSYA 32

RESULT 15

S24253

Ig heavy chain V region (VH26-DLR4-JH6) - human

C;Species: Homo sapiens (man)

C;Date: 19-Feb-1994 #sequence\_revision 10-Nov-1995 #text\_change 21-Jan-2000

C;Accession: S24253

R;Stewart, A.K.; Huang, C.; Stollar, B.D.; Schwartz, R.S.

submitted to the EMBL Data Library, June 1992

A;Description: A single VH gene predominates in the rearranged and expressed human B cell repertoires.

A;Reference number: S24247

A;Accession: S24253

A;Status: preliminary

A;Molecule type: DNA  
A;Residues: 1-109 <STE>  
A;Cross-references: EMBL:X67061  
C;Superfamily: immunoglobulin V region; immunoglobulin homology  
C;Keywords: heterotetramer; immunoglobulin  
F;12-94/Domain: immunoglobulin homology <IMM>

Query Match 26.6%; Score 51; DB 2; Length 109;  
Best Local Similarity 52.2%; Pred. No. 5.7;  
Matches 12; Conservative 2; Mismatches 7; Indels 2; Gaps 1;

Qy 1 LGTFWG--DTLNCWMLSAFSRYA 21  
||| || :|| | : || ||  
Db 7 LGTAWGVPETLLCSLWFTFSSYA 29

Search completed: January 30, 2004, 11:27:01  
Job time : 2.79767 secs

OM protein - protein search, using sw model

Run on: January 30, 2004, 11:26:28 ; Search time 3.72374 Seconds  
(without alignments)  
1841.751 Million cell updates/sec

Title: US-09-989-481-4  
Perfect score: 192  
Sequence: 1 LGTFWGDTLNCWMLSAFSRYARCLAEGHDGPTQ 33

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 789580 seqs, 207824079 residues

Total number of hits satisfying chosen parameters: 789580

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published Applications\_AA:\*

- 1: /cgn2\_6/ptodata/2/pubpaa/US07\_PUBCOMB.pep:\*
- 2: /cgn2\_6/ptodata/2/pubpaa/PCT\_NEW\_PUB.pep:\*
- 3: /cgn2\_6/ptodata/2/pubpaa/US06\_NEW\_PUB.pep:\*
- 4: /cgn2\_6/ptodata/2/pubpaa/US06\_PUBCOMB.pep:\*
- 5: /cgn2\_6/ptodata/2/pubpaa/US07\_NEW\_PUB.pep:\*
- 6: /cgn2\_6/ptodata/2/pubpaa/PCTUS\_PUBCOMB.pep:\*
- 7: /cgn2\_6/ptodata/2/pubpaa/US08\_NEW\_PUB.pep:\*
- 8: /cgn2\_6/ptodata/2/pubpaa/US08\_PUBCOMB.pep:\*
- 9: /cgn2\_6/ptodata/2/pubpaa/US09A\_PUBCOMB.pep:\*
- 10: /cgn2\_6/ptodata/2/pubpaa/US09B\_PUBCOMB.pep:\*
- 11: /cgn2\_6/ptodata/2/pubpaa/US09C\_PUBCOMB.pep:\*
- 12: /cgn2\_6/ptodata/2/pubpaa/US09\_NEW\_PUB.pep:\*
- 13: /cgn2\_6/ptodata/2/pubpaa/US10A\_PUBCOMB.pep:\*
- 14: /cgn2\_6/ptodata/2/pubpaa/US10B\_PUBCOMB.pep:\*
- 15: /cgn2\_6/ptodata/2/pubpaa/US10C\_PUBCOMB.pep:\*
- 16: /cgn2\_6/ptodata/2/pubpaa/US10\_NEW\_PUB.pep:\*
- 17: /cgn2\_6/ptodata/2/pubpaa/US60\_NEW\_PUB.pep:\*
- 18: /cgn2\_6/ptodata/2/pubpaa/US60\_PUBCOMB.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result	Query	
No.	Score	Match Length DB ID Description

1	56.5	29.4	328	11	US-09-765-061B-75	Sequence 75, Appl
2	55	28.6	306	15	US-10-081-872-319	Sequence 319, App
3	53.5	27.9	328	11	US-09-765-061B-76	Sequence 76, Appl
4	53.5	27.9	372	11	US-09-765-061B-74	Sequence 74, Appl
5	53.5	27.9	372	11	US-09-765-061B-78	Sequence 78, Appl
6	53.5	27.9	384	11	US-09-765-061B-72	Sequence 72, Appl
7	53.5	27.9	384	11	US-09-765-061B-73	Sequence 73, Appl
8	53.5	27.9	392	11	US-09-765-061B-77	Sequence 77, Appl
9	52	27.1	460	12	US-10-369-493-3584	Sequence 3584, Ap
10	51	26.6	56	9	US-09-908-711-83	Sequence 83, Appl
11	51	26.6	56	11	US-09-764-891-3081	Sequence 3081, Ap
12	51	26.6	474	9	US-09-815-242-10270	Sequence 10270, A
13	51	26.6	474	12	US-10-369-493-23518	Sequence 23518, A
14	51	26.6	1207	12	US-10-131-985-19	Sequence 19, Appl
15	50.5	26.3	390	12	US-10-214-446-10	Sequence 10, Appl
16	50.5	26.3	608	12	US-10-369-493-20224	Sequence 20224, A
17	49.5	25.8	408	12	US-10-369-493-2222	Sequence 2222, Ap
18	49.5	25.8	435	12	US-10-369-493-3999	Sequence 3999, Ap
19	49.5	25.8	842	12	US-10-190-435-2	Sequence 2, Appli
20	49.5	25.8	842	12	US-10-241-009-2	Sequence 2, Appli
21	49.5	25.8	842	12	US-10-190-434B-2	Sequence 2, Appli
22	49.5	25.8	842	12	US-10-190-305A-2	Sequence 2, Appli
23	49.5	25.8	847	10	US-09-476-242-2	Sequence 2, Appli
24	48.5	25.3	432	15	US-10-156-761-8664	Sequence 8664, Ap
25	48.5	25.3	497	15	US-10-156-761-9214	Sequence 9214, Ap
26	48	25.0	113	10	US-09-851-138-80	Sequence 80, Appl
27	48	25.0	278	12	US-10-369-493-19291	Sequence 19291, A
28	48	25.0	526	15	US-10-210-296-102	Sequence 102, App
29	48	25.0	528	12	US-10-369-493-3275	Sequence 3275, Ap
30	48	25.0	1940	14	US-10-016-283-34	Sequence 34, Appl
31	47.5	24.7	84	9	US-09-864-761-39145	Sequence 39145, A
32	47.5	24.7	414	15	US-10-156-761-9276	Sequence 9276, Ap
33	47.5	24.7	467	9	US-09-815-242-11786	Sequence 11786, A
34	47.5	24.7	865	14	US-10-055-364-24	Sequence 24, Appl
35	47	24.5	42	12	US-09-833-245-633	Sequence 633, App
36	47	24.5	42	12	US-09-833-245-635	Sequence 635, App
37	47	24.5	94	12	US-10-264-049-3360	Sequence 3360, Ap
38	46.5	24.2	107	9	US-09-864-761-41036	Sequence 41036, A
39	46.5	24.2	112	12	US-10-419-296-17	Sequence 17, Appl
40	46.5	24.2	243	12	US-10-289-776-14	Sequence 14, Appl
41	46.5	24.2	384	10	US-09-738-626-3606	Sequence 3606, Ap
42	46.5	24.2	396	15	US-10-204-887-88	Sequence 88, Appl
43	46.5	24.2	421	9	US-09-815-242-12845	Sequence 12845, A
44	46.5	24.2	421	9	US-09-932-474-3	Sequence 3, Appli
45	46.5	24.2	716	11	US-09-866-050A-183	Sequence 183, App

#### ALIGNMENTS

RESULT 1

US-09-765-061B-75

; Sequence 75, Application US/09765061B

; Publication No. US20030022165A1

; GENERAL INFORMATION:

; APPLICANT: Board of Regents of the University of Texas System

; TITLE OF INVENTION: Mutations in a No. US20030022165A1e1 Photoreceptor-pineal gene 17P cause  
 ; TITLE OF INVENTION: leber congenital amaurosis (LCA4)  
 ; FILE REFERENCE: 96606/16UTL  
 ; CURRENT APPLICATION NUMBER: US/09/765,061B  
 ; CURRENT FILING DATE: 2001-01-17  
 ; NUMBER OF SEQ ID NOS: 78  
 ; SOFTWARE: PatentIn version 3.1  
 ; SEQ ID NO 75  
 ; LENGTH: 328  
 ; TYPE: PRT  
 ; ORGANISM: Bos taurus  
 ; FEATURE:  
 ; NAME/KEY: PEPTIDE  
 ; LOCATION: (1)..(328)  
 ; OTHER INFORMATION: Cow AIPL1 Protein  
 US-09-765-061B-75

Query Match 29.4%; Score 56.5; DB 11; Length 328;  
 Best Local Similarity 43.3%; Pred. No. 11;  
 Matches 13; Conservative 5; Mismatches 11; Indels 1; Gaps 1;

QY 4 FWGDTLNCWMLSAFSRYARCLAEHGDGPTQ 33  
 || ||:: : || | :||| | ||:  
 Db 87 FWCDTIHTGVYPILSRSLRQMAEGKD-PTE 115

## RESULT 2

US-10-081-872-319  
 ; Sequence 319, Application US/10081872  
 ; Publication No. US20030125534A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Callen, Walter  
 ; APPLICANT: Richardson, Toby  
 ; APPLICANT: Frey, Gerhard  
 ; APPLICANT: Short, Jay M.  
 ; APPLICANT: Mathur, Eric J.  
 ; APPLICANT: Gray, Kevin A.  
 ; APPLICANT: Kerovuo, Janne S.  
 ; APPLICANT: Slupska, Malgorzata  
 ; TITLE OF INVENTION: ENZYMES HAVING ALPHA AMYLASE ACTIVITY  
 ; TITLE OF INVENTION: AND METHODS OF USE THEREOF  
 ; FILE REFERENCE: 09010-108001  
 ; CURRENT APPLICATION NUMBER: US/10/081,872  
 ; CURRENT FILING DATE: 2002-02-21  
 ; PRIOR APPLICATION NUMBER: US 60/270,495  
 ; PRIOR FILING DATE: 2001-02-21  
 ; PRIOR APPLICATION NUMBER: US 60/270,496  
 ; PRIOR FILING DATE: 2001-02-21  
 ; PRIOR APPLICATION NUMBER: US 60/291,122  
 ; PRIOR FILING DATE: 2001-05-14  
 ; NUMBER OF SEQ ID NOS: 321  
 ; SOFTWARE: FastSEQ for Windows Version 4.0  
 ; SEQ ID NO 319  
 ; LENGTH: 306  
 ; TYPE: PRT  
 ; ORGANISM: Artificial Sequence

; FEATURE:  
; OTHER INFORMATION: consensus sequence  
US-10-081-872-319

Query Match 28.6%; Score 55; DB 15; Length 306;  
Best Local Similarity 32.1%; Pred. No. 16;  
Matches 9; Conservative 8; Mismatches 11; Indels 0; Gaps 0;

Qy 2 GTFWGDTLNCWMLSAFSRYARCLAEAGHD 29  
||| | :: | : | :: | | : : |  
Db 108 GTFGGPDIHQWLWSSYAAYLRSIGDWFD 135

RESULT 3

US-09-765-061B-76  
; Sequence 76, Application US/09765061B  
; Publication No. US20030022165A1  
; GENERAL INFORMATION:  
; APPLICANT: Board of Regents of the University of Texas System  
; TITLE OF INVENTION: Mutations in a No. US20030022165A1el Photoreceptor-pineal  
gene 17P cause  
; TITLE OF INVENTION: leber congenital amaurosis (LCA4)  
; FILE REFERENCE: 96606/16UTL  
; CURRENT APPLICATION NUMBER: US/09/765,061B  
; CURRENT FILING DATE: 2001-01-17  
; NUMBER OF SEQ ID NOS: 78  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 76  
; LENGTH: 328  
; TYPE: PRT  
; ORGANISM: Mus musculus  
; FEATURE:  
; NAME/KEY: PEPTIDE  
; LOCATION: (1)..(328)  
; OTHER INFORMATION: Mouse AIPL1 Protein  
US-09-765-061B-76

Query Match 27.9%; Score 53.5; DB 11; Length 328;  
Best Local Similarity 44.8%; Pred. No. 28;  
Matches 13; Conservative 4; Mismatches 11; Indels 1; Gaps 1;

Qy 4 FWGDTLNCWMLSAFSRYARCLAEAGHDGPT 32  
|| ||:: : || | :||| | ||  
Db 87 FWCDTIHTGVYPMLSRSLRQVAEGKD-PT 114

RESULT 4

US-09-765-061B-74  
; Sequence 74, Application US/09765061B  
; Publication No. US20030022165A1  
; GENERAL INFORMATION:  
; APPLICANT: Board of Regents of the University of Texas System  
; TITLE OF INVENTION: Mutations in a No. US20030022165A1el Photoreceptor-pineal  
gene 17P cause  
; TITLE OF INVENTION: leber congenital amaurosis (LCA4)  
; FILE REFERENCE: 96606/16UTL  
; CURRENT APPLICATION NUMBER: US/09/765,061B

; CURRENT FILING DATE: 2001-01-17  
; NUMBER OF SEQ ID NOS: 78  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 74  
; LENGTH: 372  
; TYPE: PRT  
; ORGANISM: Papio anubis  
; FEATURE:  
; NAME/KEY: PEPTIDE  
; LOCATION: (1)..(372)  
; OTHER INFORMATION: Baboon AIPL1 Protein  
US-09-765-061B-74

Query Match 27.9%; Score 53.5; DB 11; Length 372;  
Best Local Similarity 40.0%; Pred. No. 32;  
Matches 12; Conservative 6; Mismatches 11; Indels 1; Gaps 1;

Qy 4 FWGDTLNCWMLSAFSRYARCLAEHGDGPTQ 33  
|| ||:: : || | :| | ||:  
Db 87 FWCDTIHTGVYPILSRSLRQMAQGKD-PTE 115

RESULT 5

US-09-765-061B-78

; Sequence 78, Application US/09765061B  
; Publication No. US20030022165A1  
; GENERAL INFORMATION:  
; APPLICANT: Board of Regents of the University of Texas System  
; TITLE OF INVENTION: Mutations in a No. US20030022165A1el Photoreceptor-pineal  
gene 17P cause  
; TITLE OF INVENTION: leber congenital amaurosis (LCA4)  
; FILE REFERENCE: 96606/16UTL  
; CURRENT APPLICATION NUMBER: US/09/765,061B  
; CURRENT FILING DATE: 2001-01-17  
; NUMBER OF SEQ ID NOS: 78  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 78  
; LENGTH: 372  
; TYPE: PRT  
; ORGANISM: Saimiri sciureus  
; FEATURE:  
; NAME/KEY: PEPTIDE  
; LOCATION: (1)..(372)  
; OTHER INFORMATION: Squirrel Monkey AIPL1 Protein  
US-09-765-061B-78

Query Match 27.9%; Score 53.5; DB 11; Length 372;  
Best Local Similarity 40.0%; Pred. No. 32;  
Matches 12; Conservative 6; Mismatches 11; Indels 1; Gaps 1;

Qy 4 FWGDTLNCWMLSAFSRYARCLAEHGDGPTQ 33  
|| ||:: : || | :| | ||:  
Db 87 FWCDTIHTGVYPILSRSLRQMAQGKD-PTE 115

RESULT 6

US-09-765-061B-72



```
; Sequence 72, Application US/09765061B
; Publication No. US20030022165A1
; GENERAL INFORMATION:
; APPLICANT: Board of Regents of the University of Texas System
; TITLE OF INVENTION: Mutations in a No. US20030022165A1el Photoreceptor-pineal
gene 17P cause
; TITLE OF INVENTION: leber congenital amaurosis (LCA4)
; FILE REFERENCE: 96606/16UTL
; CURRENT APPLICATION NUMBER: US/09/765,061B
; CURRENT FILING DATE: 2001-01-17
; NUMBER OF SEQ ID NOS: 78
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 72
; LENGTH: 384
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: PEPTIDE
; LOCATION: (1)..(384)
; OTHER INFORMATION: Human AIPL1 Protein
; NAME/KEY: misc_feature
; LOCATION: (322)..(322)
; OTHER INFORMATION: Xaa represents any of the twenty amino acids
US-09-765-061B-72
```

```
Query Match          27.9%; Score 53.5; DB 11; Length 384;
Best Local Similarity 40.0%; Pred. No. 33;
Matches 12; Conservative 6; Mismatches 11; Indels 1; Gaps 1;
```

```
Qy      4 FWGDTLNCWMLSAFSRYARCLAEGHDGPTQ 33
        || ||:: : || |::| | ||:
Db      87 FWCDTIHTGVYPILSRSLRQMAQGKD-PTE 115
```

#### RESULT 7

```
US-09-765-061B-73
; Sequence 73, Application US/09765061B
; Publication No. US20030022165A1
; GENERAL INFORMATION:
; APPLICANT: Board of Regents of the University of Texas System
; TITLE OF INVENTION: Mutations in a No. US20030022165A1el Photoreceptor-pineal
gene 17P cause
; TITLE OF INVENTION: leber congenital amaurosis (LCA4)
; FILE REFERENCE: 96606/16UTL
; CURRENT APPLICATION NUMBER: US/09/765,061B
; CURRENT FILING DATE: 2001-01-17
; NUMBER OF SEQ ID NOS: 78
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 73
; LENGTH: 384
; TYPE: PRT
; ORGANISM: Pan troglodytes
; FEATURE:
; NAME/KEY: PEPTIDE
; LOCATION: (1)..(384)
; OTHER INFORMATION: Chimpanzee AIPL1 Protein
US-09-765-061B-73
```

Query Match 27.9%; Score 53.5; DB 11; Length 384;  
Best Local Similarity 40.0%; Pred. No. 33;  
Matches 12; Conservative 6; Mismatches 11; Indels 1; Gaps 1;

Qy 4 FWGDTLNCWMLSAFSRYARCLAEHGDGPTQ 33  
|| ||:: : || | :|:| | ||:  
Db 87 FWCDTIHTGVYPILSRSLRQMAQGKD-PTE 115

RESULT 8

US-09-765-061B-77

; Sequence 77, Application US/09765061B  
; Publication No. US20030022165A1  
; GENERAL INFORMATION:  
; APPLICANT: Board of Regents of the University of Texas System  
; TITLE OF INVENTION: Mutations in a No. US20030022165A1el Photoreceptor-pineal  
gene 17P cause  
; TITLE OF INVENTION: leber congenital amaurosis (LCA4)  
; FILE REFERENCE: 96606/16UTL  
; CURRENT APPLICATION NUMBER: US/09/765,061B  
; CURRENT FILING DATE: 2001-01-17  
; NUMBER OF SEQ ID NOS: 78  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 77  
; LENGTH: 392  
; TYPE: PRT  
; ORGANISM: Macaca mulatta  
; FEATURE:  
; NAME/KEY: PEPTIDE  
; LOCATION: (1)..(392)  
; OTHER INFORMATION: Rhesus Monkey AIPL1 Protein  
US-09-765-061B-77

Query Match 27.9%; Score 53.5; DB 11; Length 392;  
Best Local Similarity 40.0%; Pred. No. 33;  
Matches 12; Conservative 6; Mismatches 11; Indels 1; Gaps 1;

Qy 4 FWGDTLNCWMLSAFSRYARCLAEHGDGPTQ 33  
|| ||:: : || | :|:| | ||:  
Db 87 FWCDTIHTGVYPILSRSLRQMAQGKD-PTE 115

RESULT 9

US-10-369-493-3584

; Sequence 3584, Application US/10369493  
; Publication No. US20030233675A1  
; GENERAL INFORMATION:  
; APPLICANT: Cao, Yongwei  
; APPLICANT: Hinkle, Gregory J.  
; APPLICANT: Slater, Steven C.  
; APPLICANT: Goldman, Barry S.  
; APPLICANT: Chen, Xianfeng  
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION  
OF  
; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES  
; FILE REFERENCE: 38-10(52052)B

```
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 3584
; LENGTH: 460
; TYPE: PRT
; ORGANISM: Neurospora crassa
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(460)
; OTHER INFORMATION: unsure at all Xaa locations
US-10-369-493-3584
```

```
Query Match          27.1%; Score 52; DB 12; Length 460;
Best Local Similarity 50.0%; Pred. No. 63;
Matches 10; Conservative 3; Mismatches 7; Indels 0; Gaps 0;
```

```
Qy      10 NCWMLSAFSRYARCLAEGHD 29
        | |:| | | | | : | |
Db      430 NSWVLSAKSDYRRIVSSGQD 449
```

RESULT 10

US-09-908-711-83

```
; Sequence 83, Application US/09908711
; Patent No. US20020045230A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PA128
; CURRENT APPLICATION NUMBER: US/09/908,711
; CURRENT FILING DATE: 2001-07-20
; PRIOR APPLICATION NUMBER: US01/01360
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: 09/764,867
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: US01/01344
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: 09/764,892
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: US01/01345
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: 09/764,888
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: US01/01329
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: 09/764,905
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: US01/01354
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: 09/764,891
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: US01/01339
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: 09/764,869
```

; PRIOR FILING DATE: 2001-01-17  
; PRIOR APPLICATION NUMBER: US01/01340  
; PRIOR FILING DATE: 2001-01-17  
; PRIOR APPLICATION NUMBER: 09/764,874  
; PRIOR FILING DATE: 2001-01-17  
; PRIOR APPLICATION NUMBER: US01/01334  
; PRIOR FILING DATE: 2001-01-17  
; PRIOR APPLICATION NUMBER: 09/764,898  
; PRIOR FILING DATE: 2001-01-17  
; PRIOR APPLICATION NUMBER: US01/01320  
; PRIOR FILING DATE: 2001-01-17  
; PRIOR APPLICATION NUMBER: 09/764,853  
; PRIOR FILING DATE: 2001-01-17  
; PRIOR APPLICATION NUMBER: US01/01349  
; PRIOR FILING DATE: 2001-01-17  
; PRIOR APPLICATION NUMBER: 09/764,902  
; PRIOR FILING DATE: 2001-01-17  
; PRIOR APPLICATION NUMBER: US01/01239  
; PRIOR FILING DATE: 2001-01-17  
; PRIOR APPLICATION NUMBER: 09/764,870  
; PRIOR FILING DATE: 2001-01-17  
; PRIOR APPLICATION NUMBER: US01/01348  
; PRIOR FILING DATE: 2001-01-17  
; PRIOR APPLICATION NUMBER: 09/764,882  
; PRIOR FILING DATE: 2001-01-17  
; PRIOR APPLICATION NUMBER: US01/01347  
; PRIOR FILING DATE: 2001-01-17  
; PRIOR APPLICATION NUMBER: 09/764,896  
; PRIOR FILING DATE: 2001-01-17  
; PRIOR APPLICATION NUMBER: US01/01307  
; PRIOR FILING DATE: 2001-01-17  
; PRIOR APPLICATION NUMBER: 09/764,864  
; PRIOR FILING DATE: 2001-01-17  
; PRIOR APPLICATION NUMBER: US01/01341  
; PRIOR FILING DATE: 2001-01-17  
; PRIOR APPLICATION NUMBER: 09/764,856  
; PRIOR FILING DATE: 2001-01-17  
; PRIOR APPLICATION NUMBER: US01/01336  
; PRIOR FILING DATE: 2001-01-17  
; PRIOR APPLICATION NUMBER: 09/764,868  
; PRIOR FILING DATE: 2001-01-17  
; PRIOR APPLICATION NUMBER: US01/01312  
; PRIOR FILING DATE: 2001-01-17  
; PRIOR APPLICATION NUMBER: 60/179,065  
; PRIOR FILING DATE: 2000-01-31  
; PRIOR APPLICATION NUMBER: 60/180,628  
; PRIOR FILING DATE: 2000-02-04  
; PRIOR APPLICATION NUMBER: 60/209,467  
; PRIOR FILING DATE: 2000-06-07  
; NUMBER OF SEQ ID NOS: 167  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 83  
; LENGTH: 56  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: SITE

; LOCATION: (12)  
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
US-09-908-711-83

Query Match 26.6%; Score 51; DB 9; Length 56;  
Best Local Similarity 57.9%; Pred. No. 10;  
Matches 11; Conservative 1; Mismatches 5; Indels 2; Gaps 1;

Qy 9 LNCWMLSAFSRYAR--CLA 25  
|||| || |: | |||  
Db 25 LNCWHLSCFNHALRLSCLA 43

RESULT 11

US-09-764-891-3081  
; Sequence 3081, Application US/09764891  
; Publication No. US20030077808A1  
; GENERAL INFORMATION:  
; APPLICANT: Rosen et al.  
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
; FILE REFERENCE: PC006  
; CURRENT APPLICATION NUMBER: US/09/764,891  
; CURRENT FILING DATE: 2001-01-17  
; Prior application data removed - consult PALM or file wrapper  
; NUMBER OF SEQ ID NOS: 10231  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 3081  
; LENGTH: 56  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: SITE  
; LOCATION: (12)  
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
US-09-764-891-3081

Query Match 26.6%; Score 51; DB 11; Length 56;  
Best Local Similarity 57.9%; Pred. No. 10;  
Matches 11; Conservative 1; Mismatches 5; Indels 2; Gaps 1;

Qy 9 LNCWMLSAFSRYAR--CLA 25  
|||| || |: | |||  
Db 25 LNCWHLSCFNHALRLSCLA 43

RESULT 12

US-09-815-242-10270  
; Sequence 10270, Application US/09815242  
; Patent No. US20020061569A1  
; GENERAL INFORMATION:  
; APPLICANT: Haselbeck, Robert  
; APPLICANT: Ohlsen, Kari L.  
; APPLICANT: Zyskind, Judith W.  
; APPLICANT: Wall, Daniel  
; APPLICANT: Trawick, John D.  
; APPLICANT: Carr, Grant J.  
; APPLICANT: Yamamoto, Robert T.

; APPLICANT: Xu, H. Howard  
 ; TITLE OF INVENTION: Identification of Essential Genes in  
 ; TITLE OF INVENTION: Prokaryotes  
 ; FILE REFERENCE: ELITRA.011A  
 ; CURRENT APPLICATION NUMBER: US/09/815,242  
 ; CURRENT FILING DATE: 2001-03-21  
 ; PRIOR APPLICATION NUMBER: 60/191,078  
 ; PRIOR FILING DATE: 2000-03-21  
 ; PRIOR APPLICATION NUMBER: 60/206,848  
 ; PRIOR FILING DATE: 2000-05-23  
 ; PRIOR APPLICATION NUMBER: 60/207,727  
 ; PRIOR FILING DATE: 2000-05-26  
 ; PRIOR APPLICATION NUMBER: 60/242,578  
 ; PRIOR FILING DATE: 2000-10-23  
 ; PRIOR APPLICATION NUMBER: 60/253,625  
 ; PRIOR FILING DATE: 2000-11-27  
 ; PRIOR APPLICATION NUMBER: 60/257,931  
 ; PRIOR FILING DATE: 2000-12-22  
 ; PRIOR APPLICATION NUMBER: 60/269,308  
 ; PRIOR FILING DATE: 2001-02-16  
 ; NUMBER OF SEQ ID NOS: 14110  
 ; SOFTWARE: FastSEQ for Windows Version 4.0  
 ; SEQ ID NO 10270  
 ; LENGTH: 474  
 ; TYPE: PRT  
 ; ORGANISM: Escherichia coli  
 US-09-815-242-10270

Query Match 26.6%; Score 51; DB 9; Length 474;  
 Best Local Similarity 43.3%; Pred. No. 89;  
 Matches 13; Conservative 3; Mismatches 14; Indels 0; Gaps 0;

Qy 1 LGTFWGDTLNCWMLSAFSRYARCLAEGHDG 30  
 | | : | | : | | | | | | | |  
 Db 141 LVTEYGSWRNRKLVEFFSRYARTCFEAFDG 170

# RESULT 13

US-10-369-493-23518

; Sequence 23518, Application US/10369493

; Publication No. US20030233675A1

## ; GENERAL INFORMATION:

; APPLICANT: Cao, Yongwei

; APPLICANT: Hinkle, Gregory J.

; APPLICANT: Slater, Steven C.

; APPLICANT: Goldman, Barry S.

; APPLICANT: Chen, Xianfeng

; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF

; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES

; FILE REFERENCE: 38-10(52052)B

; CURRENT APPLICATION NUMBER: US/10/369,493

; CURRENT FILING DATE: 2003-02-28

; PRIOR APPLICATION NUMBER: US 60/360,039

; PRIOR FILING DATE: 2002-02-21

; NUMBER OF SEQ ID NOS: 47374

; SEQ ID NO 23518

; LENGTH: 474  
; TYPE: PRT  
; ORGANISM: Escherichia coli  
US-10-369-493-23518

Query Match 26.6%; Score 51; DB 12; Length 474;  
Best Local Similarity 43.3%; Pred. No. 89;  
Matches 13; Conservative 3; Mismatches 14; Indels 0; Gaps 0;

Qy 1 LGTFWGDTLNCWMLSAFSRYARCLAEAGHDG 30  
| | : | | :: ||||| | ||  
Db 141 LVTEYGSWRNRKLVFFSRYARTCFEAFDG 170

RESULT 14

US-10-131-985-19  
; Sequence 19, Application US/10131985  
; Publication No. US20030199440A1  
; GENERAL INFORMATION:  
; APPLICANT: Dack, Kevin N  
; APPLICANT: Davies, Michael J  
; APPLICANT: Fish, Paul V  
; APPLICANT: Huggins, Jonathan P  
; APPLICANT: McIntosh, Fraser S  
; APPLICANT: Occleston, Nicholas L  
; TITLE OF INVENTION: Composition  
; FILE REFERENCE: PCS 10391A  
; CURRENT APPLICATION NUMBER: US/10/131,985  
; CURRENT FILING DATE: 2002-04-25  
; PRIOR APPLICATION NUMBER: US/09/726,295  
; PRIOR FILING DATE: 2000-11-30  
; PRIOR APPLICATION NUMBER: GB 9930768.8  
; PRIOR FILING DATE: 1999-12-29  
; NUMBER OF SEQ ID NOS: 60  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 19  
; LENGTH: 1207  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-131-985-19

Query Match 26.6%; Score 51; DB 12; Length 1207;  
Best Local Similarity 56.2%; Pred. No. 2.3e+02;  
Matches 9; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

Qy 18 SRYARCLAEAGHDGPTQ 33  
| ||||: || | |  
Db 841 SMYARCISEGEDATCQ 856

RESULT 15

US-10-214-446-10  
; Sequence 10, Application US/10214446  
; Publication No. US20030180742A1  
; GENERAL INFORMATION:  
; APPLICANT: Weiner, David  
; APPLICANT: Burk, Mark J.

```
; APPLICANT: Hitchman, Tim
; APPLICANT: Pujol, Catherine
; APPLICANT: Richardson, Toby
; APPLICANT: Short, Jay M.
; TITLE OF INVENTION: P450 ENZYMES, NUCLEIC ACIDS ENCODING
; TITLE OF INVENTION: THEM AND METHODS OF MAKING AND USING THEM
; FILE REFERENCE: 09010-500001
; CURRENT APPLICATION NUMBER: US/10/214,446
; CURRENT FILING DATE: 2002-08-05
; PRIOR APPLICATION NUMBER: US 60/309,497
; PRIOR FILING DATE: 2001-08-03
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 10
; LENGTH: 390
; TYPE: PRT
; ORGANISM: Bacterial
US-10-214-446-10
```

```
Query Match          26.3%; Score 50.5; DB 12; Length 390;
Best Local Similarity 45.5%; Pred. No. 85;
Matches    10; Conservative    4; Mismatches    5; Indels    3; Gaps    1;
```

```
QY          4 FWGDTLNCWMLSAFSRYARCLA 25
             || : | |:|  :|| |||
Db          30 FWHELLGSWVL---TRHADCLA 48
```

```
Search completed: January 30, 2004, 11:35:27
Job time : 3.72374 secs
```



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 30, 2004, 11:17:27 ; Search time 3.91634 Seconds  
(without alignments)  
2174.410 Million cell updates/sec

Title: US-09-989-481-4  
Perfect score: 192  
Sequence: 1 LGTFWGDTLNCWMLSAFSRYARCLAEGHDGPTQ 33

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SPTREMBL\_23:\*  
1: sp\_archaea:\*  
2: sp\_bacteria:\*  
3: sp\_fungi:\*  
4: sp\_human:\*  
5: sp\_invertebrate:\*  
6: sp\_mammal:\*  
7: sp\_mhc:\*  
8: sp\_organelle:\*  
9: sp\_phage:\*  
10: sp\_plant:\*  
11: sp\_rodent:\*  
12: sp\_virus:\*  
13: sp\_vertebrate:\*  
14: sp\_unclassified:\*  
15: sp\_rvirus:\*  
16: sp\_bacteriaphage:\*  
17: sp\_archaeophages:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

%  
Result Query  
No. Score Match Length DB ID Description  
-----

1	56.5	29.4	111	5	Q96868	Q96868 heliocidari
2	56.5	29.4	111	5	Q96869	Q96869 heliocidari
3	56.5	29.4	196	16	Q8VJS3	Q8vjs3 mycobacteri
4	56.5	29.4	223	16	Q10843	Q10843 mycobacteri
5	56.5	29.4	328	6	Q95MP1	Q95mp1 bos taurus
6	56.5	29.4	565	2	Q9ZNN9	Q9znn9 comamonas t
7	56.5	29.4	584	2	Q9S150	Q9s150 comamonas t
8	56.5	29.4	680	5	Q8WSN8	Q8wsn8 caenorhabdi
9	55	28.6	515	8	Q8SME7	Q8sme7 globba plat
10	55	28.6	515	8	Q8HV78	Q8hv78 cornukaempf
11	55	28.6	694	10	Q9XFS2	Q9xfs2 arabidopsis
12	54.5	28.4	191	6	Q9N2C2	Q9n2c2 oryctolagus
13	54	28.1	287	3	Q9UT16	Q9ut16 schizosacch
14	54	28.1	316	16	Q9AAP0	Q9aap0 caulobacter
15	53.5	27.9	179	11	Q8R057	Q8r057 mus musculu
16	53.5	27.9	328	11	Q924K1	Q924k1 mus musculu
17	53.5	27.9	372	6	Q95MN7	Q95mn7 saimiri bol
18	53.5	27.9	372	6	Q95MN8	Q95mn8 papio cynoc
19	53.5	27.9	384	6	Q95MN9	Q95mn9 pan paniscu
20	53.5	27.9	392	6	Q95MP0	Q95mp0 macaca mula
21	53	27.6	144	4	Q13051	Q13051 homo sapien
22	53	27.6	511	8	Q8HV20	Q8hv20 siphonochil
23	53	27.6	782	10	Q9AUV9	Q9auv9 oryza sativ
24	52.5	27.3	952	2	Q9FB53	Q9fb53 corynebacte
25	52	27.1	255	16	Q9A321	Q9a321 caulobacter
26	52	27.1	379	2	Q9AGI5	Q9agi5 pseudomonas
27	52	27.1	412	8	Q8HV03	Q8hv03 orchidantha
28	52	27.1	515	8	Q8HV41	Q8hv41 paramomum p
29	52	27.1	515	8	Q8HV32	Q8hv32 renealmia c
30	52	27.1	533	10	Q81510	Q81510 arabidopsis
31	52	27.1	859	15	Q8UTD6	Q8utd6 human immun
32	51.5	26.8	1150	3	Q99129	Q99129 ustilago ma
33	51	26.6	355	8	Q9GHQ8	Q9ghq8 persea indi
34	51	26.6	355	8	Q9GHX8	Q9ghx8 endlicheria
35	51	26.6	355	8	Q9GI00	Q9gi00 cassytha ci
36	51	26.6	355	8	Q9GHQ5	Q9ghq5 persea ling
37	51	26.6	373	10	Q9M0B8	Q9m0b8 arabidopsis
38	51	26.6	452	8	Q8HV02	Q8hv02 phenakosper
39	51	26.6	472	16	Q9RVZ5	Q9rvz5 deinococcus
40	51	26.6	474	16	Q8X841	Q8x841 escherichia
41	51	26.6	483	8	Q8MA80	Q8ma80 brunia albi
42	51	26.6	511	8	Q8HV19	Q8hv19 siphonochil
43	51	26.6	511	8	Q8HV18	Q8hv18 siphonochil
44	51	26.6	511	8	Q8HV04	Q8hv04 musella las
45	51	26.6	513	8	Q8WKE4	Q8wke4 hedychium s

#### ALIGNMENTS

##### RESULT 1

Q96868

ID Q96868 PRELIMINARY; PRT; 111 AA.

AC Q96868;

DT 01-MAY-1999 (TrEMBLrel. 10, Created)

DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)

DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)

DE Cell signaling molecule Wnt-5 (Fragment).  
 GN HEWNT-5.  
 OS Heliocidaris erythrogramma (Sea urchin).  
 OC Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;  
 OC Echinoidea; Euechinoidea; Echinacea; Echinoida; Echinometridae;  
 OC Heliocidaris.  
 OX NCBI\_TaxID=7634;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=98320638; PubMed=9656482;  
 RA Ferkowicz M.J., Stander M.C., Raff R.A.;  
 RT "Phylogenetic relationships and developmental expression of three sea  
 RT urchin Wnt genes."  
 RL Mol. Biol. Evol. 15:809-819(1998).  
 DR EMBL; U58983; AAC69434.1; -.  
 DR InterPro; IPR005817; Wnt.  
 DR Pfam; PF00110; wnt; 1.  
 DR SMART; SM00097; WNT1; 1.  
 FT NON\_TER 1 1  
 FT NON\_TER 111 111  
 SQ SEQUENCE 111 AA; 12353 MW; 2E33E31CE69A05AD CRC64;

Query Match 29.4%; Score 56.5; DB 5; Length 111;  
 Best Local Similarity 50.0%; Pred. No. 2.1;  
 Matches 12; Conservative 3; Mismatches 8; Indels 1; Gaps 1;

Qy 11 CWM-LSAFSRYARCLAEGHDGPTQ 33  
 ||: || |:| | | :|| ||  
 Db 5 CWLQLSPFNRVGSILKEKYDGATQ 28

## RESULT 2

O96869

ID O96869 PRELIMINARY; PRT; 111 AA.  
 AC O96869;  
 DT 01-MAY-1999 (TrEMBLrel. 10, Created)  
 DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)  
 DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)  
 DE Cell signaling molecule Wnt-5 (Fragment).  
 GN HTWNT-5.  
 OS Heliocidaris tuberculata (Sea urchin).  
 OC Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;  
 OC Echinoidea; Euechinoidea; Echinacea; Echinoida; Echinometridae;  
 OC Heliocidaris.  
 OX NCBI\_TaxID=7635;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=98320638; PubMed=9656482;  
 RA Ferkowicz M.J., Stander M.C., Raff R.A.;  
 RT "Phylogenetic relationships and developmental expression of three sea  
 RT urchin Wnt genes."  
 RL Mol. Biol. Evol. 15:809-819(1998).  
 DR EMBL; U58984; AAC69435.1; -.  
 DR InterPro; IPR005817; Wnt.  
 DR Pfam; PF00110; wnt; 1.  
 DR SMART; SM00097; WNT1; 1.  
 FT NON\_TER 1 1

FT NON\_TER 111 111  
SQ SEQUENCE 111 AA; 12341 MW; A858F5718F388D4D CRC64;

Query Match 29.4%; Score 56.5; DB 5; Length 111;  
Best Local Similarity 50.0%; Pred. No. 2.1;  
Matches 12; Conservative 3; Mismatches 8; Indels 1; Gaps 1;

Qy 11 CWM-LSAFSRYARCLAEGHDGPTQ 33  
||: || |:| | | :|| ||  
Db 5 CWLQLSPFN RVGSILKEKYDGATQ 28

RESULT 3

Q8VJS3

ID Q8VJS3 PRELIMINARY; PRT; 196 AA.  
AC Q8VJS3;  
DT 01-MAR-2002 (TrEMBLrel. 20, Created)  
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)  
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)  
DE IS1607, transposase.  
GN MT2070.  
OS Mycobacterium tuberculosis.  
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.  
OX NCBI\_TaxID=1773;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CDC 1551 / Oshkosh;  
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,  
RA Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,  
RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,  
RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,  
RA Bishai W.;  
RT "Whole genome comparison of Mycobacterium tuberculosis clinical and  
RT laboratory strains."  
RL Submitted (APR-2001) to the EMBL/GenBank/DDBJ databases.  
DR EMBL; AE007058; AAK46348.1; -.  
DR TIGR; MT2070; -.  
DR InterPro; IPR003346; Transposase\_20.  
DR Pfam; PF02371; Transposase\_20; 1.  
SQ SEQUENCE 196 AA; 21349 MW; C145A8D836FD9C2D CRC64;

Query Match 29.4%; Score 56.5; DB 16; Length 196;  
Best Local Similarity 46.7%; Pred. No. 3.7;  
Matches 14; Conservative 2; Mismatches 9; Indels 5; Gaps 2;

Qy 4 FWGDT--LNCWMLSAFSRYARCLAEGHDGP 31  
| ||: | | | || |:| ||| |  
Db 126 FAGDSRRANLW---AADRYNRAIARGHDHP 152

RESULT 4

Q10843

ID Q10843 PRELIMINARY; PRT; 223 AA.  
AC Q10843;  
DT 01-NOV-1998 (TrEMBLrel. 08, Created)  
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)

DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)  
 DE Hypothetical protein Rv2014.  
 GN RV2014 OR MTCY39.03C.  
 OS Mycobacterium tuberculosis.  
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
 OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.  
 OX NCBI\_TaxID=1773;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=H37Rv;  
 RX MEDLINE=98295987; PubMed=9634230;  
 RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,  
 RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekaia F.,  
 RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,  
 RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,  
 RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,  
 RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,  
 RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,  
 RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;  
 RT "Deciphering the biology of Mycobacterium tuberculosis from the  
 RT complete genome sequence."  
 RL Nature 393:537-544(1998).  
 CC -!- SIMILARITY: TO M.PARATUBERCULOSIS IS900.  
 DR EMBL; Z74025; CAA98415.1; -.  
 DR TubercuList; Rv2014; -.  
 DR InterPro; IPR003346; Transposase\_20.  
 DR Pfam; PF02371; Transposase\_20; 1.  
 KW Hypothetical protein; Complete proteome.  
 SQ SEQUENCE 223 AA; 24132 MW; 70456750017FEF37 CRC64;

Query Match 29.4%; Score 56.5; DB 16; Length 223;  
 Best Local Similarity 46.7%; Pred. No. 4.2;  
 Matches 14; Conservative 2; Mismatches 9; Indels 5; Gaps 2;

Qy 4 FWGDT--LNCWMLSAFSRYARCLAEGHDGP 31  
 | ||: | | | || | :| ||| |  
 Db 157 FAGDSRRANLW---AADRYNRAIARGHDHP 183

# RESULT 5

## Q95MP1

ID Q95MP1 PRELIMINARY; PRT; 328 AA.  
 AC Q95MP1;  
 DT 01-DEC-2001 (TrEMBLrel. 19, Created)  
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)  
 DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)  
 DE Aryl-hydrocarbon interacting protein-like 1.  
 GN AIPL1.  
 OS Bos taurus (Bovine).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;  
 OC Bovidae; Bovinae; Bos.  
 OX NCBI\_TaxID=9913;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=21313649; PubMed=11420621;  
 RA Sohocki M.M., Sullivan L.S., Tirpak D.L., Daiger S.P.;

RT "Comparative analysis of aryl-hydrocarbon receptor interacting  
 RT protein-like 1 (Aipl1), a gene associated with inherited retinal  
 RT disease in humans.";  
 RL Mamm. Genome 12:566-568(2001).  
 DR EMBL; AF296410; AAK77954.1; -.  
 DR InterPro; IPR001440; TPR.  
 DR Pfam; PF00515; TPR; 2.  
 SQ SEQUENCE 328 AA; 38472 MW; B2B5E7ACF5E0A72A CRC64;

Query Match 29.4%; Score 56.5; DB 6; Length 328;  
 Best Local Similarity 43.3%; Pred. No. 6.3;  
 Matches 13; Conservative 5; Mismatches 11; Indels 1; Gaps 1;

Qy 4 FWGDTLNCWMLSAFSRYARCLAEAGHDGPTQ 33  
 || ||:: : || | :||| | ||:  
 Db 87 FWCDTIHTGVYPILSRSLRQMAEGKD-PTE 115

# RESULT 6

Q9ZNN9

ID Q9ZNN9 PRELIMINARY; PRT; 565 AA.  
 AC Q9ZNN9;  
 DT 01-MAY-1999 (TrEMBLrel. 10, Created)  
 DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
 DE AphR protein.  
 GN APHR.  
 OS Comamonas testosteroni (Pseudomonas testosteroni).  
 OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;  
 OC Comamonadaceae; Comamonas.  
 OX NCBI\_TaxID=285;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=TA441;  
 RX MEDLINE=99018839; PubMed=9802031;  
 RA Arai H., Akahira S., Ohishi T., Maeda M., Kudo T.;  
 RT "Adaptation of Comamonas testosteroni TA441 to utilize phenol:  
 RT organization and regulation of the genes involved in phenol  
 RT degradation.";  
 RL Microbiology 144:2895-2903(1998).  
 CC -!- SIMILARITY: CONTAINS A SIGMA-54 FACTOR INTERACTION ATP-BINDING  
 CC DOMAIN.  
 DR EMBL; AB006480; BAA34177.1; -.  
 DR InterPro; IPR003593; AAA\_ATPase.  
 DR InterPro; IPR002197; HTH\_Fis.  
 DR InterPro; IPR002078; Sig54\_interact.  
 DR InterPro; IPR004096; V4R.  
 DR Pfam; PF02954; HTH\_8; 1.  
 DR Pfam; PF00158; Sigma54\_activat; 1.  
 DR Pfam; PF02830; V4R; 1.  
 DR PRINTS; PR01590; HTHFIS.  
 DR SMART; SM00382; AAA; 1.  
 DR TIGRFAMs; TIGR01199; HTH\_fis; 1.  
 DR PROSITE; PS00675; SIGMA54\_INTERACT\_1; 1.  
 DR PROSITE; PS00676; SIGMA54\_INTERACT\_2; 1.  
 DR PROSITE; PS00688; SIGMA54\_INTERACT\_3; 1.  
 DR PROSITE; PS50045; SIGMA54\_INTERACT\_4; 1.

KW ATP-binding; DNA-binding; Transcription; Transcription regulation.  
SQ SEQUENCE 565 AA; 62649 MW; D6D0F0AD984D3201 CRC64;

Query Match 29.4%; Score 56.5; DB 2; Length 565;  
Best Local Similarity 33.3%; Pred. No. 11;  
Matches 18; Conservative 3; Mismatches 6; Indels 27; Gaps 4;

QY 5 WG--DTLNCWML-----SAFSR-----YARCLAEG---HDGP 31  
|| | :||| || | :| || || |: |  
Db 138 WGPQDQPSCWMLLGYSAGYSSAFFRRPVFFKEMQCSTCGHAHCLIEGRFQHEWP 191

RESULT 7

Q9S150

ID Q9S150 PRELIMINARY; PRT; 584 AA.  
AC Q9S150;  
DT 01-MAY-2000 (TrEMBLrel. 13, Created)  
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)  
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
DE Positive regulator of phenol-degradative genes.  
GN PHCR.  
OS Comamonas testosteroni (Pseudomonas testosteroni).  
OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;  
OC Comamonadaceae; Comamonas.  
OX NCBI\_TaxID=285;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=R5;  
RX MEDLINE=20055761; PubMed=10589844;  
RA Teramoto M., Futamata H., Harayama S., Watanabe K.;  
RT "Characterization of a high-affinity phenol hydroxylase from Comamonas  
RT testosteroni R5 by gene cloning, and expression in Pseudomonas  
RT aeruginosa PA01c.";  
RL Mol. Gen. Genet. 262:552-558(1999).  
CC -!- SIMILARITY: CONTAINS A SIGMA-54 FACTOR INTERACTION ATP-BINDING  
CC DOMAIN.  
DR EMBL; AB024741; BAA87867.1; -.  
DR InterPro; IPR003593; AAA\_ATPase.  
DR InterPro; IPR002197; HTH\_Fis.  
DR InterPro; IPR002078; Sig54\_interact.  
DR InterPro; IPR004096; V4R.  
DR Pfam; PF02954; HTH\_8; 1.  
DR Pfam; PF00158; Sigma54\_activat; 1.  
DR Pfam; PF02830; V4R; 1.  
DR PRINTS; PR01590; HTHFIS.  
DR SMART; SM00382; AAA; 1.  
DR TIGRFAMs; TIGR01199; HTH\_fis; 1.  
DR PROSITE; PS00675; SIGMA54\_INTERACT\_1; 1.  
DR PROSITE; PS00676; SIGMA54\_INTERACT\_2; 1.  
DR PROSITE; PS00688; SIGMA54\_INTERACT\_3; 1.  
DR PROSITE; PS50045; SIGMA54\_INTERACT\_4; 1.  
KW ATP-binding; DNA-binding; Transcription; Transcription regulation.  
SQ SEQUENCE 584 AA; 64666 MW; 94AB4D5612513158 CRC64;

Query Match 29.4%; Score 56.5; DB 2; Length 584;  
Best Local Similarity 33.3%; Pred. No. 11;  
Matches 18; Conservative 3; Mismatches 6; Indels 27; Gaps 4;

QY 5 WG--DTLNCWML-----SAFSR-----YARCLAEG---HDGP 31  
 || | :||| ||| | :| || || |: |  
 Db 155 WGPQDQPSCWMLLG YASGYSSAFFRRPVFFKEMQCSTCGHAHCLIEGRFQHEWP 208

RESULT 8

Q8WSN8

ID Q8WSN8 PRELIMINARY; PRT; 680 AA.  
 AC Q8WSN8;  
 DT 01-MAR-2002 (TrEMBLrel. 20, Created)  
 DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
 DE Hypothetical 77.4 kDa protein.  
 GN Y41D4B.26.  
 OS Caenorhabditis elegans.  
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  
 OC Rhabditidae; Peloderinae; Caenorhabditis.  
 OX NCBI\_TaxID=6239;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Bristol N2;  
 RX MEDLINE=99069613; PubMed=9851916;  
 RA None;  
 RT "Genome sequence of the nematode C. elegans: a platform for  
 RT investigating biology. The C. elegans Sequencing Consortium."  
 RL Science 282:2012-2018(1998).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Bristol N2;  
 RA Geisel C., Lamar B.;  
 RT "The sequence of C. elegans cosmid Y41D4B."  
 RL Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Bristol N2;  
 RA Waterston R.;  
 RT "Direct Submission."  
 RL Submitted (DEC-2001) to the EMBL/GenBank/DDBJ databases.  
 CC -!- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).  
 CC -!- SIMILARITY: BELONGS TO THE NUCLEAR HORMONE RECEPTOR FAMILY.  
 DR EMBL; AC024776; AAL32241.1; -.  
 DR WormPep; Y41D4B.26; CE30003.  
 DR InterPro; IPR000536; Hormone\_rec\_lig.  
 DR InterPro; IPR001628; Znf\_C4steroid.  
 DR Pfam; PF00104; hormone\_rec; 1.  
 DR Pfam; PF00105; zf-C4; 1.  
 DR PRINTS; PR00047; STROIDFINGER.  
 DR ProDom; PD000035; Znf\_C4steroid; 1.  
 DR SMART; SM00430; HOLI; 1.  
 DR SMART; SM00399; Znf\_C4; 1.  
 KW Hypothetical protein; DNA-binding; Metal-binding; Nuclear protein;  
 KW Receptor; Transcription; Transcription regulation; Zinc; Zinc-finger.  
 SQ SEQUENCE 680 AA; 77412 MW; F870B89A2C162305 CRC64;

Query Match 29.4%; Score 56.5; DB 5; Length 680;  
 Best Local Similarity 43.3%; Pred. No. 13;



Matches 13; Conservative 5; Mismatches 5; Indels 7; Gaps 2;

Qy 5 WGDTLNCWML----SAF---SRYARCLAEG 27  
||: :|| : ||| |:|:|||| |  
Db 42 WGEFVNCCEIVSTGSAFCKSCRFKCLAVG 71

RESULT 9

Q8SME7

ID Q8SME7 PRELIMINARY; PRT; 515 AA.  
AC Q8SME7;  
DT 01-JUN-2002 (TrEMBLrel. 21, Created)  
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)  
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)  
DE Intron maturase (Maturase K).  
GN MATK.  
OS Globba platystachya.  
OG Chloroplast.  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; Liliopsida; Zingiberales; Zingiberaceae;  
OC Globba.  
OX NCBI\_TaxID=138161;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Takano A., Okada H.;  
RT "Multiple occurrences of triploid formation in Globba (Zingiberaceae)  
RT from molecular evidence.";  
RL Plant Syst. Evol. 230:143-159(2002).  
CC -!- FUNCTION: PROBABLY ASSISTS IN SPLICING CHLOROPLAST GROUP II  
CC INTRONS (BY SIMILARITY).  
CC -!- SIMILARITY: WITH CORRESPONDING ORF IN OTHER PLANT CHLOROPLASTS,  
CC AND REGIONS OF SIMILARITY TO MATURASE-LIKE POLYPEPTIDES ENCODED BY  
CC MITOCHONDRIAL INTRONS.  
DR EMBL; AB049250; BAB85874.1; -.  
DR InterPro; IPR000442; Intron\_maturase2.  
DR InterPro; IPR002866; MatK\_N.  
DR Pfam; PF01348; Intron\_maturas2; 1.  
DR Pfam; PF01824; MatK\_N; 1.  
KW mRNA processing; Chloroplast.  
SQ SEQUENCE 515 AA; 62022 MW; DE784AD0C3F48B5A CRC64;

Query Match 28.6%; Score 55; DB 8; Length 515;  
Best Local Similarity 34.5%; Pred. No. 16;  
Matches 10; Conservative 7; Mismatches 12; Indels 0; Gaps 0;

Qy 5 WGDTLNCWMLSAFSRYARCLAEGHDGPTQ 33  
| | :| ::| || | |: | | ::  
Db 398 WTDLADCDIINRFSRICRKLSHYHSGSSK 426

RESULT 10

Q8HV78

ID Q8HV78 PRELIMINARY; PRT; 515 AA.  
AC Q8HV78;  
DT 01-MAR-2003 (TrEMBLrel. 23, Created)  
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)  
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)

DE Maturase K.  
 GN MATK.  
 OS Cornukaempferia aurantiflora.  
 OG Chloroplast.  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; Liliopsida; Zingiberales; Zingiberaceae;  
 OC Cornukaempferia.  
 OX NCBI\_TaxID=97739;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Leaf;  
 RA Kress W.J., Prince L.M., Williams K.J.;  
 RT "The phylogeny and a new classification of the gingers  
 RT (Zingiberaceae): Evidence from molecular data.";  
 RL Am. J. Bot. 89:1684-1698(2002).  
 DR EMBL; AF478835; AAN63192.1; -.  
 KW Chloroplast.  
 SQ SEQUENCE 515 AA; 62125 MW; 63FB8C35B66CEA29 CRC64;

Query Match 28.6%; Score 55; DB 8; Length 515;  
 Best Local Similarity 34.5%; Pred. No. 16;  
 Matches 10; Conservative 7; Mismatches 12; Indels 0; Gaps 0;

QY 5 WGDTLNCWMLSAFSRYARCLAEGHDGPTQ 33  
 | | : | :: | | | | : | | ::  
 Db 398 WTDLADCDIINRFSRICRKLSHYHSGSSK 426

# RESULT 11

## Q9XFS2

ID Q9XFS2 PRELIMINARY; PRT; 694 AA.  
 AC Q9XFS2;  
 DT 01-NOV-1999 (TrEMBLrel. 12, Created)  
 DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)  
 DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)  
 DE Cyclic nucleotide and calmodulin-regulated ion channel  
 DE (AT5G54250/MDK4\_7).  
 GN CNGC4.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
 OX NCBI\_TaxID=3702;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=99272993; PubMed=10341447;  
 RA Kohler C., Merkle T., Neuhaus G.;  
 RT "Characterisation of a novel gene family of putative cyclic  
 RT nucleotide-and calmodulin-regulated ion channels in Arabidopsis  
 RT thaliana.";  
 RL Plant J. 18:97-104(1999).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA Koehler C.;  
 RL Submitted (AUG-1998) to the EMBL/GenBank/DDBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.

RC STRAIN=Columbia;  
 RX MEDLINE=98344145; PubMed=9679202;  
 RA Kaneko T., Kotani H., Nakamura Y., Sato S., Asamizu E., Miyajima N.,  
 RA Tabata S.;  
 RT "Structural analysis of Arabidopsis thaliana chromosome 5. V. Sequence  
 RT features of the regions of 1,381,565 bp covered by twenty one  
 RT physically assigned P1 and TAC clones.";  
 RL DNA Res. 5:131-145(1998).  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RA Cheuk R., Chen H., Kim C.J., Koesema E., Meyers M.C., Banh J.,  
 RA Bowser L., Carninci P., Dale J.M., Goldsmith A.D., Hayashizaki Y.,  
 RA Ishida J., Jiang P.X., Jones T., Kamiya A., Karlin-Neumann G.,  
 RA Kawai J., Lam B., Lee J.M., Lin J., Liu S.X., Miranda M., Narusaka M.,  
 RA Nguyen M., Onodera C.S., Palm C.J., Pham P.K., Quach H.L., Sakurai T.,  
 RA Satou M., Seki M., Southwick A., Tang C.C., Toriumi M., Yamada K.,  
 RA Yamamura Y., Yu G., Yu S., Shinozaki K., Davis R.W., Theologis A.,  
 RA Ecker J.R.;  
 RT "Arabidopsis cDNA clones.";  
 RL Submitted (SEP-2001) to the EMBL/GenBank/DDBJ databases.  
 DR EMBL; Y17912; CAB40129.1; -.  
 DR EMBL; AB010695; BAB10748.1; -.  
 DR EMBL; AY057691; AAL15321.1; -.  
 DR InterPro; IPR000595; cNMP\_binding.  
 DR InterPro; IPR005821; Ion\_trans.  
 DR Pfam; PF00027; cNMP\_binding; 1.  
 DR Pfam; PF00520; ion\_trans; 1.  
 DR SMART; SM00100; cNMP; 1.  
 DR PROSITE; PS50042; CNMP\_BINDING\_3; 1.  
 KW Ionic channel; Transmembrane.  
 SQ SEQUENCE 694 AA; 80081 MW; E3F843AE1B0F1EA0 CRC64;

Query Match 28.6%; Score 55; DB 10; Length 694;  
 Best Local Similarity 35.9%; Pred. No. 22;  
 Matches 14; Conservative 2; Mismatches 9; Indels 14; Gaps 2;

QY 2 GT-FWGDTLN-----CWMLSAFSRYARCLAE 26  
 || :|| || || | | :|| |  
 Db 255 GTVWWGIALNMIAYFVAHAAGACWYLLGVQRSAKCLKE 293

# RESULT 12

Q9N2C2

ID Q9N2C2 PRELIMINARY; PRT; 191 AA.  
 AC Q9N2C2;  
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)  
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
 DE Prostaglandin D synthase.  
 OS Oryctolagus cuniculus (Rabbit).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.  
 OX NCBI\_TaxID=9986;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Brain;  
 RA Nakau H., Fujimori K., Urade Y.;

RT "Isolation of rabbit cDNA for lipocalin-type prostaglandin D  
RT synthase.";  
RL Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.  
DR EMBL; AB040991; BAA94343.1; -.  
DR HSSP; P80188; 1DFV.  
DR InterPro; IPR002345; Lipocalin.  
DR InterPro; IPR000566; Lipocln\_cytFABP.  
DR Pfam; PF00061; lipocalin; 1.  
DR PRINTS; PR00179; LIPOCALIN.  
DR PROSITE; PS00213; LIPOCALIN; 1.  
SQ SEQUENCE 191 AA; 21444 MW; 1424BD9878512F61 CRC64;

Query Match 28.4%; Score 54.5; DB 6; Length 191;  
Best Local Similarity 33.3%; Pred. No. 7;  
Matches 10; Conservative 6; Mismatches 13; Indels 1; Gaps 1;

QY 5 WGDTLNCWMLSA-FSRYARCLAEHGDGPTQ 33  
|| | : |:: : :| :|| || |  
Db 112 WGSTYSVWVVDTDYKEFALLYSEGAKGPGQ 141

RESULT 13

Q9UT16

ID Q9UT16 PRELIMINARY; PRT; 287 AA.  
AC Q9UT16;  
DT 01-MAY-2000 (TrEMBLrel. 13, Created)  
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)  
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
DE Yeast atp12 protein precursor homolog.  
GN SPAC9.12C.  
OS Schizosaccharomyces pombe (Fission yeast).  
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;  
OC Schizosaccharomycetales; Schizosaccharomycetaceae;  
OC Schizosaccharomyces.  
OX NCBI\_TaxID=4896;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=972h-;  
RA Wedler H., Duesterhoeft A., Lyne M.H., Rajandream M.A., Barrell B.G.;  
RL Submitted (OCT-1999) to the EMBL/GenBank/DDBJ databases.  
DR EMBL; AL121764; CAB57430.1; -.  
DR GeneDB\_SPombe; SPAC9.12c; -.  
SQ SEQUENCE 287 AA; 33149 MW; 21F78CCD7B2FFD97 CRC64;

Query Match 28.1%; Score 54; DB 3; Length 287;  
Best Local Similarity 47.4%; Pred. No. 12;  
Matches 9; Conservative 2; Mismatches 8; Indels 0; Gaps 0;

Qy 5 WGDTLNCWMLSAFSRYARC 23  
| :|| | |:|| | |  
Db 198 WLSSLNSWQLAAFERSVSC 216

RESULT 14

Q9AAP0

ID Q9AAP0 PRELIMINARY; PRT; 316 AA.  
AC Q9AAP0;

DT 01-JUN-2001 (TrEMBLrel. 17, Created)  
 DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)  
 DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)  
 DE Hypothetical protein CC0557.  
 GN CC0557.  
 OS Caulobacter crescentus.  
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Caulobacterales;  
 OC Caulobacteraceae; Caulobacter.  
 OX NCBI\_TaxID=155892;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=ATCC 19089 / CB15;  
 RX MEDLINE=21173698; PubMed=11259647;  
 RA Nierman W.C., Feldblyum T.V., Laub M.T., Paulsen I.T., Nelson K.E.,  
 RA Eisen J., Heidelberg J.F., Alley M.R.K., Ohta N., Maddock J.R.,  
 RA Potocka I., Nelson W.C., Newton A., Stephens C., Phadke N.D., Ely B.,  
 RA DeBoy R.T., Dodson R.J., Durkin A.S., Gwinn M.L., Haft D.H.,  
 RA Kolonay J.F., Smit J., Craven M.B., Khouri H., Shetty J., Berry K.,  
 RA Utterback T., Tran K., Wolf A., Vamathevan J., Ermolaeva M., White O.,  
 RA Salzberg S.L., Venter J.C., Shapiro L., Fraser C.M.;  
 RT "Complete genome sequence of *Caulobacter crescentus*.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 98:4136-4141(2001).  
 DR EMBL; AE005729; AAK22543.1; -.  
 DR TIGR; CC0557; -.  
 KW Hypothetical protein; Complete proteome.  
 SQ SEQUENCE 316 AA; 35026 MW; 41C4289216FED963 CRC64;

Query Match 28.1%; Score 54; DB 16; Length 316;  
 Best Local Similarity 56.2%; Pred. No. 14;  
 Matches 9; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

Qy 6 GDTLNCWMLSAFSRYA 21  
 || |:|| | | ||:  
 Db 301 GDILSCWKLGAVPRYS 316

#### RESULT 15

Q8R057

ID Q8R057 PRELIMINARY; PRT; 179 AA.  
 AC Q8R057;  
 DT 01-JUN-2002 (TrEMBLrel. 21, Created)  
 DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)  
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)  
 DE Similar to aryl-hydrocarbon interacting protein-like 1.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Eye;  
 RA Strausberg R.;  
 RL Submitted (APR-2002) to the EMBL/GenBank/DDBJ databases.  
 DR EMBL; BC028285; AAH28285.1; -.  
 SQ SEQUENCE 179 AA; 20424 MW; 32ED79C343761A10 CRC64;

Query Match 27.9%; Score 53.5; DB 11; Length 179;

Best Local Similarity 44.8%; Pred. No. 9.1;  
Matches 13; Conservative 4; Mismatches 11; Indels 1; Gaps 1;

Qy 4 FWGDTLNCWMLSAFSRYARCLAEGHDGPT 32  
|||::: : || |:| | ||  
Db 87 FWCDTIHTGVYPMLSRSLRQVAEGKD-PT 114

Search completed: January 30, 2004, 11:26:22  
Job time : 4.91634 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 30, 2004, 11:15:32 ; Search time 1.28405 Seconds  
(without alignments)  
1208.586 Million cell updates/sec

Title: US-09-989-481-4  
Perfect score: 192  
Sequence: 1 LGTFWGD<sup>T</sup>LNCWMLSAFSRYARCLAE<sup>G</sup>HDGPTQ 33

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SwissProt\_41:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query		DB	ID	Description
		Match	Length			
1	59	30.7	910	1	IMB2_SCHPO	O14089 schizosacch
2	53.5	27.9	328	1	AIPL_RAT	Q9jlg9 rattus norv
3	53.5	27.9	384	1	AIPL_HUMAN	Q9nzn9 homo sapien
4	52	27.1	563	1	LIP1_GEOCN	P17573 geotrichum
5	51	26.6	418	1	CGA1_XENLA	P18606 xenopus lae
6	51	26.6	474	1	ASCB_ECOLI	P24240 escherichia
7	51	26.6	1207	1	EGF_HUMAN	P01133 homo sapien
8	50	26.0	504	1	MATK_EICCR	Q9ghb1 eichhornia
9	49.5	25.8	355	1	GBA2_NEUCR	Q05424 neurospora
10	49.5	25.8	847	1	ENV_HV1S1	P19550 human immun
11	49	25.5	282	1	APAH_BURMA	Q9aev8 burkholderi
12	49	25.5	282	1	APAH_BURPS	O69115 burkholderi
13	48	25.0	229	1	PEPE_ECOL6	Q8fb55 escherichia
14	48	25.0	229	1	PEPE_ECOLI	P32666 escherichia
15	48	25.0	472	1	HEAD_BP <sup>G</sup> A1	Q9fzw7 bacterioph
16	48	25.0	503	1	MATK_PSINU	Q8wi35 psilotum nu
17	48	25.0	512	1	MATK_ACECA	Q8sm90 acer campes

18	48	25.0	517	1	MATK_ACEPS	Q8se90	acer pseudo
19	48	25.0	1959	1	AGRI_RAT	P25304	rattus norv
20	47	24.5	690	1	PPK_PSEAE	Q9s646	pseudomonas
21	47	24.5	1513	1	MUC2_RAT	Q62635	rattus norv
22	47	24.5	3433	1	POLG_KUNJM	P14335	k genome po
23	46.5	24.2	729	1	NARB_SYNP7	P39458	synechococc
24	46.5	24.2	895	1	ODP1_ALCEU	Q59097	alcaligenes
25	46	24.0	182	1	C560_CAEEL	P41956	caenorhabdi
26	46	24.0	184	1	C560_CAEBR	P41955	caenorhabdi
27	46	24.0	361	1	COOH_RHORU	P31895	rhodospiril
28	46	24.0	449	1	HEAD_BPB03	Q37888	bacterioph
29	46	24.0	512	1	MATK_LILHE	Q9gih9	lilium henr
30	46	24.0	512	1	MATK_LILRE	Q9ghc3	lilium rega
31	46	24.0	1550	1	GLTB_SYNY3	P55037	synechocyst
32	46	24.0	1822	1	ITB4_HUMAN	P16144	homo sapien
33	45.5	23.7	126	1	YF81_XYLFA	Q9p9t2	xylella fas
34	45.5	23.7	614	1	VAA1_DROME	P48602	drosophila
35	45	23.4	114	1	RSN_MOUSE	Q99p87	mus musculu
36	45	23.4	158	1	NEU4_ONCKE	P16042	oncorhynch
37	45	23.4	334	1	GBLP_ORYSA	P49027	oryza sativ
38	45	23.4	404	1	VE2_HP60	Q80944	human papil
39	45	23.4	698	1	PPK_XYLFA	Q9pac7	xylella fas
40	45	23.4	1597	1	SOL_DROME	P27398	drosophila
41	45	23.4	3038	1	TRIO_HUMAN	O75962	homo sapien
42	45	23.4	3430	1	POLG_WNV	P06935	w genome po
43	44.5	23.2	191	1	PGHD_FELCA	Q29487	felis silve
44	44.5	23.2	276	1	PLPB_PASHA	Q08869	pasteurella
45	44.5	23.2	500	1	GABT_BOVIN	Q9bgi0	bos taurus

# ALIGNMENTS

## RESULT 1

### IMB2\_SCHPO

ID IMB2\_SCHPO STANDARD; PRT; 910 AA..

AC O14089;

DT 15-DEC-1998 (Rel. 37, Created)

DT 15-DEC-1998 (Rel. 37, Last sequence update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE Putative importin beta-2 subunit (Karyopherin beta-2 subunit)

DE (Importin 104) (Transportin) (TRN).

GN SPAC2F3.06C.

OS Schizosaccharomyces pombe (Fission yeast).

OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;

OC Schizosaccharomycetales; Schizosaccharomycetaceae;

OC Schizosaccharomyces.

OX NCBI\_TaxID=4896;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=972;

RX MEDLINE=21848401; PubMed=11859360;

RA Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,

RA Sgouros J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,

RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,

RA Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,

RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,



RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,  
 RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,  
 RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,  
 RA Oliver K., O'Neil S., Pearson D., Quail M.A., Rabbino-witsch E.,  
 RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,  
 RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,  
 RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,  
 RA Woodward J., Volckaert G., Aert R., Robben J., Grymonprez B.,  
 RA Weltjens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,  
 RA Gabel C., Fuchs M., Fritz C., Holzer E., Moestl D., Hilbert H.,  
 RA Borzym K., Langer I., Beck A., Lehrach H., Reinhardt R., Pohl T.M.,  
 RA Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,  
 RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,  
 RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,  
 RA Lucas M., Rochet M., Gaillardin C., Tallada V.A., Garzon A., Thode G.,  
 RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,  
 RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,  
 RA Cerrutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,  
 RA Shpakovski G.V., Ussery D., Barrell B.G., Nurse P.;  
 RT "The genome sequence of *Schizosaccharomyces pombe*.";  
 RL Nature 415:871-880(2002).  
 CC -!- FUNCTION: REQUIRED FOR IMPORT OF MRNA BINDING PROTEINS. BINDS TO  
 CC NUCLEOPORINS (BY SIMILARITY).  
 CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).  
 CC -!- SIMILARITY: BELONGS TO THE IMPORTIN BETA FAMILY.  
 CC -!- SIMILARITY: Contains 1 importin N-terminal domain.  
 CC -!- SIMILARITY: Contains 9 HEAT repeats.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; Z99165; CAB16272.1; -.  
 DR PIR; T38539; T38539.  
 DR HSSP; Q92973; 1QBK.  
 DR GeneDB\_SPombe; SPAC2F3.06c; -.  
 DR InterPro; IPR000357; HEAT\_repeat.  
 DR InterPro; IPR001494; Importinb\_N.  
 DR PROSITE; PS50077; HEAT\_REPEAT; FALSE\_NEG.  
 DR PROSITE; PS50166; IMPORTIN\_B\_NT; FALSE\_NEG.  
 KW Hypothetical protein; Transport; Protein transport; Repeat.  
 FT DOMAIN 34 122 IMPORTIN N-TERMINAL.  
 FT REPEAT 127 164 HEAT 1.  
 FT REPEAT 174 211 HEAT 2.  
 FT REPEAT 299 336 HEAT 3.  
 FT REPEAT 410 447 HEAT 4.  
 FT REPEAT 451 488 HEAT 5.  
 FT REPEAT 497 534 HEAT 6.  
 FT REPEAT 538 575 HEAT 7.  
 FT REPEAT 769 808 HEAT 8.  
 FT REPEAT 850 890 HEAT 9.  
 FT DOMAIN 366 385 ASP/GLU-RICH (ACIDIC).  
 SQ SEQUENCE 910 AA; 101718 MW; 4939CD9B09877208 CRC64;

Query Match 30.7%; Score 59; DB 1; Length 910;  
Best Local Similarity 40.9%; Pred. No. 1.7;  
Matches 9; Conservative 4; Mismatches 9; Indels 0; Gaps 0;

Qy 8 TLNCWMLSAFSRYARCLAEGHD 29  
|: || | :|::| || |  
Db 473 TITCWTLGGRYSKWASCLESEED 494

RESULT 2

AIPL\_RAT

ID AIPL\_RAT STANDARD; PRT; 328 AA.  
AC Q9JLG9;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Aryl-hydrocarbon interacting protein-like 1.  
GN AIPL1.  
OS Rattus norvegicus (Rat).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
OX NCBI\_TaxID=10116;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=20082814; PubMed=10615133;

RA Sohocki M.M., Bowne S.J., Sullivan L.S., Blackshaw S., Cepko C.L.,

RA Payne A.M., Bhattacharya S.S., Khaliq S., Mehdi Q., Birch D.G.,

RA Harrison W.R., Elder F.F.B., Heckenlively J.R., Daiger S.P.;

RT "Mutations in a novel photoreceptor-pineal gene on 17p cause Leber

RT congenital amaurosis.";

RL Nat. Genet. 24:79-83(2000).

CC -!- TISSUE SPECIFICITY: HIGHLY EXPRESSED IN RETINA.

CC -!- SIMILARITY: BELONGS TO THE FKBP-TYPE PPIASE FAMILY.

CC -!- SIMILARITY: Contains 2 TPR repeats.

-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----

DR EMBL; AF180340; AAF26707.1; -.

DR InterPro; IPR001179; FKBP\_PPIase.

DR InterPro; IPR001440; TPR.

DR Pfam; PF00515; TPR; 2.

DR PROSITE; PS00453; FKBP\_PPIASE\_1; FALSE\_NEG.

DR PROSITE; PS00454; FKBP\_PPIASE\_2; FALSE\_NEG.

DR PROSITE; PS50059; FKBP\_PPIASE\_3; FALSE\_NEG.

KW Repeat; TPR repeat.

FT DOMAIN 53 145 PPIASE, FKBP-TYPE.

FT REPEAT 230 263 TPR 1.

FT REPEAT 264 297 TPR 2.

SQ SEQUENCE 328 AA; 38294 MW; E9BC3A4084F64A0E CRC64;

Query Match 27.9%; Score 53.5; DB 1; Length 328;  
 Best Local Similarity 44.8%; Pred. No. 3.5;  
 Matches 13; Conservative 4; Mismatches 11; Indels 1; Gaps 1;

Qy 4 FWGDTLNCWMLSAFSRYARCLAEGHDGPT 32  
 || ||:: : || | :||| | ||  
 Db 87 FWCDTIHTGVYPMLSRSLRQVAEGKD-PT 114

### RESULT 3

#### AIPL\_HUMAN

ID AIPL HUMAN STANDARD; PRT; 384 AA.  
 AC Q9NZN9; Q9H873; Q9NS10;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 15-SEP-2003 (Rel. 42, Last annotation update)  
 DE Aryl-hydrocarbon interacting protein-like 1.  
 GN AIPL1.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A., VARIANT HIS-90, AND VARIANT LCA4 ARG-239.  
 RX MEDLINE=20082814; PubMed=10615133;  
 RA Sohocki M.M., Bowne S.J., Sullivan L.S., Blackshaw S., Cepko C.L.,  
 RA Payne A.M., Bhattacharya S.S., Khaliq S., Mehdi Q., Birch D.G.,  
 RA Harrison W.R., Elder F.F.B., Heckenlively J.R., Daiger S.P.;  
 RT "Mutations in a novel photoreceptor-pineal gene on 17p cause Leber  
 RT congenital amaurosis.";  
 RL Nat. Genet. 24:79-83(2000).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=99265969; PubMed=10331942;  
 RA Sohocki M.M., Malone K.A., Sullivan L.S., Daiger S.P.;  
 RT "Localization of retina/pineal-expressed sequences: identification of  
 RT novel candidate genes for inherited retinal disorders.";  
 RL Genomics 58:29-33(1999).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RA Isogai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,  
 RA Nishikawa T., Nagai K., Sugano S., Shiratori A., Sudo H.,  
 RA Wagatsuma M., Hosoiri T., Kaku Y., Kodaira H., Kondo H., Sugawara M.,  
 RA Takahashi M., Chiba Y., Ishida S., Murakawa K., Ono Y., Takiguchi S.,  
 RA Watanabe S., Kimura K., Murakami K., Ishii S., Kawai Y., Saito K.,  
 RA Yamamoto J., Wakamatsu A., Nakamura Y., Nagahari K., Masuho Y.,  
 RA Ninomiya K., Iwayanagi T.;  
 RT "NEDO human cDNA sequencing project.";  
 RL Submitted (AUG-2000) to the EMBL/GenBank/DDBJ databases.  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Eye;  
 RX MEDLINE=22388257; PubMed=12477932;  
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,

RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
 RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,  
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,  
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
 RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
 RA Fahey J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,  
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
 RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,  
 RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;  
 RT "Generation and initial analysis of more than 15,000 full-length  
 RT human and mouse cDNA sequences.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
 CC -!- TISSUE SPECIFICITY: HIGHLY EXPRESSED IN RETINA.  
 CC -!- DISEASE: Defects in AIPL1 are the cause of Leber congenital  
 CC amaurosis type 4 (LCA4) [MIM:604393]; a disease characterized by  
 CC total blindness or greatly impaired vision with loss of central  
 CC vision.  
 CC -!- SIMILARITY: BELONGS TO THE FKBP-TYPE PPIASE FAMILY.  
 CC -!- SIMILARITY: Contains 2 TPR repeats.  
 CC -!- DATABASE: NAME=Mutations of the AIPL1 gene.  
 CC NOTE=Retina International's Scientific Newsletter;  
 CC WWW="http://www.retina-international.com/sci-news/aip1mut.htm".  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; AF180472; AAF26708.1; -.  
 DR EMBL; AF148864; AAF74023.1; -.  
 DR EMBL; AK023970; BAB14744.1; -.  
 DR EMBL; BC012055; AAH12055.1; -.  
 DR Genew; HGNC:359; AIPL1.  
 DR MIM; 604392; -.  
 DR MIM; 604393; -.  
 DR GO; GO:0005634; C:nucleus; TAS.  
 DR GO; GO:0003754; F:chaperone activity; TAS.  
 DR GO; GO:0007601; P:vision; TAS.  
 DR InterPro; IPR001179; FKBP\_PPIase.  
 DR InterPro; IPR001440; TPR.  
 DR Pfam; PF00515; TPR; 2.  
 DR PROSITE; PS00453; FKBP\_PPIASE\_1; FALSE\_NEG.  
 DR PROSITE; PS00454; FKBP\_PPIASE\_2; FALSE\_NEG.  
 DR PROSITE; PS50059; FKBP\_PPIASE\_3; FALSE\_NEG.  
 KW Repeat; TPR repeat; Disease mutation; Vision.  
 FT DOMAIN 53 145 PPIASE, FKBP-TYPE.  
 FT REPEAT 230 263 TPR 1.  
 FT REPEAT 264 297 TPR 2.  
 FT VARIANT 90 90 D -> H.  
 FT /FTId=VAR\_010140.

FT VARIANT 239 239 C -> R (in LCA4).  
 FT /FTid=VAR\_010139.  
 FT CONFLICT 306 315 RLLLENRMAEK -> EAAGEPHGGE (IN REF. 1).  
 SQ SEQUENCE 384 AA; 43903 MW; 47F681A1DC91A82D CRC64;

Query Match 27.9%; Score 53.5; DB 1; Length 384;  
 Best Local Similarity 40.0%; Pred. No. 4.2;  
 Matches 12; Conservative 6; Mismatches 11; Indels 1; Gaps 1;

Qy 4 FWGDTLNCWMLSAFSRYARCLAEGHDGPTQ 33  
 || ||:: : || | :|| | ||:  
 Db 87 FWCDTIHTGVYPILSRSLRQMAQGKD-PTE 115

#### RESULT 4

##### LIP1\_GEOCN

ID LIP1\_GEOCN STANDARD; PRT; 563 AA.  
 AC P17573;  
 DT 01-AUG-1990 (Rel. 15, Created)  
 DT 01-AUG-1990 (Rel. 15, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Lipase 1 precursor (EC 3.1.1.3).  
 GN LIP1.  
 OS *Geotrichum candidum* (Oospora lactis).  
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 OC Saccharomycetales; Dipodascaceae; Galactomyces.  
 OX NCBI\_TaxID=27317;  
 RN [1]  
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.  
 RC STRAIN=ATCC 34614;  
 RX MEDLINE=90110016; PubMed=2481674;  
 RA Shimada Y., Sugihara A., Tominaga Y., Iizumi T., Tsunasawa S.;  
 RT "cDNA molecular cloning of *Geotrichum candidum* lipase.";  
 RL J. Biochem. 106:383-388(1989).  
 RN [2]  
 RP SIMILARITY TO CARBOXYLESTERASES.  
 RX MEDLINE=90328988; PubMed=2115773;  
 RA Slabas A.R., Windust J., Sidebottom C.M.;  
 RT "Does sequence similarity of human choline esterase, Torpedo  
 RT acetylcholine esterase and *Geotrichum candidum* lipase reveal the  
 RT active site serine residue?";  
 RL Biochem. J. 269:279-280(1990).  
 RN [3]  
 RP X-RAY CRYSTALLOGRAPHY (2.2 ANGSTROMS).  
 RX MEDLINE=91287805; PubMed=2062369;  
 RA Schrag J.D., Li Y., Wu S., Cygler M.;  
 RT "Ser-His-Glu triad forms the catalytic site of the lipase from  
 RT *Geotrichum candidum*.";  
 RL Nature 351:761-765(1991).  
 CC -!- FUNCTION: THE EXTRACELLULAR LIPASE PRODUCED BY *G.CANDIDUM*  
 CC HYDROLYZES ALL ESTER BONDS IN TRIGLYCERIDE AND DISPLAYS A HIGH  
 CC AFFINITY FOR TRIOLEIN.  
 CC -!- CATALYTIC ACTIVITY: Triacylglycerol + H(2)O = diacylglycerol + a  
 CC fatty acid anion.  
 CC -!- SUBUNIT: Monomer.  
 CC -!- SIMILARITY: Belongs to the type-B carboxylesterase/lipase family.  
 DR PIR; PN0492; ACGUGC.

DR PDB; 1THG; 31-OCT-93.  
 DR InterPro; IPR002018; CarbesteraseB.  
 DR InterPro; IPR000379; Ser\_estrs\_site.  
 DR Pfam; PF00135; COesterase; 1.  
 DR PROSITE; PS00122; CARBOXYLESTERASE\_B\_1; 1.  
 DR PROSITE; PS00941; CARBOXYLESTERASE\_B\_2; 1.  
 KW Hydrolase; Lipid degradation; Glycoprotein; Signal; 3D-structure;  
 KW Pyrrolidone carboxylic acid.  
 FT SIGNAL 1 19  
 FT CHAIN 20 563 LIPASE 1.  
 FT MOD\_RES 20 20 PYRROLIDONE CARBOXYLIC ACID.  
 FT ACT\_SITE 236 236  
 FT ACT\_SITE 373 373  
 FT ACT\_SITE 482 482  
 FT DISULFID 80 124  
 FT DISULFID 295 307  
 FT CARBOHYD 302 302 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 383 383 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT STRAND 23 26  
 FT TURN 27 29  
 FT STRAND 30 33  
 FT STRAND 35 37  
 FT TURN 38 39  
 FT STRAND 40 47  
 FT HELIX 54 56  
 FT TURN 57 58  
 FT TURN 69 70  
 FT STRAND 71 72  
 FT STRAND 74 74  
 FT STRAND 79 79  
 FT HELIX 85 96  
 FT HELIX 98 101  
 FT HELIX 104 113  
 FT TURN 114 114  
 FT STRAND 120 120  
 FT STRAND 126 132  
 FT TURN 133 134  
 FT TURN 137 138  
 FT STRAND 141 147  
 FT TURN 151 152  
 FT HELIX 156 159  
 FT HELIX 163 171  
 FT TURN 172 173  
 FT STRAND 177 181  
 FT HELIX 186 190  
 FT HELIX 194 199  
 FT TURN 200 200  
 FT TURN 202 203  
 FT HELIX 204 219  
 FT HELIX 220 223  
 FT TURN 224 224  
 FT STRAND 225 235  
 FT TURN 236 236  
 FT HELIX 237 247  
 FT HELIX 248 250  
 FT TURN 251 251  
 FT STRAND 254 255

FT	TURN	256	257
FT	STRAND	258	259
FT	STRAND	263	267
FT	HELIX	285	293
FT	TURN	294	294
FT	TURN	297	298
FT	HELIX	301	310
FT	HELIX	313	327
FT	TURN	329	331
FT	HELIX	334	336
FT	TURN	337	337
FT	HELIX	351	356
FT	TURN	357	358
FT	STRAND	365	370
FT	TURN	371	371
FT	STRAND	372	372
FT	TURN	373	377
FT	HELIX	378	381
FT	TURN	382	383
FT	HELIX	387	397
FT	TURN	398	400
FT	HELIX	403	412
FT	HELIX	417	419
FT	TURN	423	424
FT	TURN	426	429
FT	HELIX	435	446
FT	TURN	447	447
FT	HELIX	448	457
FT	TURN	459	460
FT	STRAND	463	468
FT	TURN	470	474
FT	TURN	476	478
FT	STRAND	481	481
FT	TURN	482	485
FT	HELIX	486	490
FT	TURN	491	491
FT	TURN	495	496
FT	HELIX	497	510
FT	TURN	513	514
FT	TURN	527	529
FT	STRAND	531	535
FT	STRAND	540	544
FT	TURN	547	548
FT	HELIX	549	557
FT	HELIX	559	561
SQ	SEQUENCE	563 AA;	61230 MW; 3B7327678CB7BAAA CRC64;

Query Match 27.1%; Score 52; DB 1; Length 563;  
 Best Local Similarity 40.0%; Pred. No. 9.9;  
 Matches 14; Conservative 1; Mismatches 14; Indels 6; Gaps 1;

Qy	1	LGTFWGDTL-----NCWMLSAFSRYARCLAEGHD	29
Db	478	LGTFHGSDLLFQYYAGPWSSSAYRRYFISFANHHD	512

RESULT 5

CGAL\_XENLA  
ID CGAL\_XENLA STANDARD; PRT; 418 AA.  
AC P18606;  
DT 01-NOV-1990 (Rel. 16, Created)  
DT 01-NOV-1990 (Rel. 16, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Cyclin A1.  
OS *Xenopus laevis* (African clawed frog).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidae; Pipidae;  
OC Xenopodinae; *Xenopus*.  
OX NCBI\_TaxID=8355;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Ovary;  
RX MEDLINE=90360999; PubMed=2143983;  
RA Minshull J., Golsteyn R., Hill C.S., Hunt T.;  
RT "The A- and B-type cyclin associated cdc2 kinases in *Xenopus* turn on  
RT and off at different times in the cell cycle.";  
RL EMBO J. 9:2865-2875(1990).  
CC -!- FUNCTION: MAY BE INVOLVED IN THE CONTROL OF THE CELL CYCLE AT THE  
CC G1/S (START) AND G2/M (MITOSIS) TRANSITIONS (BY SIMILARITY).  
CC -!- SUBUNIT: INTERACTS WITH THE CDK2 AND THE CDC2 PROTEIN KINASES TO  
CC FORM A SERINE/THREONINE KINASE HOLOENZYME COMPLEX. THE CYCLIN  
CC SUBUNIT IMPARTS SUBSTRATE SPECIFICITY TO THE COMPLEX (BY  
CC SIMILARITY).  
CC -!- DEVELOPMENTAL STAGE: PRESENT IN EGGS AND EARLY EMBRYOS BUT CANNOT  
CC BE DETECTED IN LATE EMBRYOS.  
CC -!- SIMILARITY: BELONGS TO THE CYCLIN FAMILY. CYCLIN AB SUBFAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; X53745; CAA37775.1; -.  
DR PIR; S11678; S11678.  
DR HSSP; P30274; 1VIN.  
DR InterPro; IPR006670; Cyclin.  
DR InterPro; IPR004367; Cyclin\_Cterm.  
DR InterPro; IPR006671; Cyclin\_N.  
DR Pfam; PF00134; cyclin; 1.  
DR Pfam; PF02984; cyclin\_C; 1.  
DR SMART; SM00385; CYCLIN; 2.  
DR PROSITE; PS00292; CYCLINS; 1.  
KW Cyclin; Cell cycle; Cell division; Mitosis.  
SQ SEQUENCE 418 AA; 46772 MW; FEA0B7A1F8011E6A CRC64;

Query Match 26.6%; Score 51; DB 1; Length 418;  
Best Local Similarity 44.0%; Pred. No. 10;  
Matches 11; Conservative 3; Mismatches 11; Indels 0; Gaps 0;

Qy 4 FWGDTLNCWMLSAFSRYARCLAEQH 28  
|| ||| : | | | ||:: |



Db

356 FWPDTLEAFTGYALSIDIAPCLSDLH 380

RESULT 6

ASCB\_ECOLI

ID ASCB\_ECOLI STANDARD; PRT; 474 AA.  
AC P24240; P78104; Q59375;  
DT 01-MAR-1992 (Rel. 21, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE 6-phospho-beta-glucosidase ascB (EC 3.2.1.86).  
GN ASCB OR B2716.  
OS Escherichia coli.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
OC Enterobacteriaceae; Escherichia.  
OX NCBI\_TaxID=562;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=K12;  
RX MEDLINE=92334140; PubMed=1630307;  
RA Hall B.G., Xu L.;  
RT "Nucleotide sequence, function, activation, and evolution of the  
RT cryptic asc operon of Escherichia coli K12.";  
RL Mol. Biol. Evol. 9:688-706(1992).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=K12 / MG1655;  
RX MEDLINE=97426617; PubMed=9278503;  
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,  
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,  
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,  
RA Mau B., Shao Y.;  
RT "The complete genome sequence of Escherichia coli K-12.";  
RL Science 277:1453-1474(1997).  
CC -!- FUNCTION: CAN HYDROLYZE SALICIN, CELLOBIOSE, AND PROBABLY  
CC ARBUTIN.  
CC -!- CATALYTIC ACTIVITY: 6-phospho-beta-D-glucoside-(1,4)-D-glucose +  
CC H(2)O = D-glucose 6-phosphate + D-glucose.  
CC -!- SIMILARITY: BELONGS TO FAMILY 1 OF GLYCOSYL HYDROLASES.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; M73326; AAA16430.1; -.  
DR EMBL; U29579; AAA69226.1; ALT\_INIT.  
DR EMBL; AE000355; AAC75758.1; -.  
DR PIR; H65051; H65051.  
DR HSSP; P11546; 1PBG.  
DR EcoGene; EG10085; ascB.  
DR InterPro; IPR001360; Glyco\_hydro\_1.  
DR Pfam; PF00232; Glyco\_hydro\_1; 1.  
DR PRINTS; PR00131; GLHYDRIASE1.

DR ProDom; PD000650; Glyco\_hydro\_1; 1.  
 DR PROSITE; PS00572; GLYCOSYL\_HYDROL\_F1\_1; 1.  
 DR PROSITE; PS00653; GLYCOSYL\_HYDROL\_F1\_2; 1.  
 KW Hydrolase; Glycosidase; Complete proteome.  
 FT ACT\_SITE 180 180 PROTON DONOR (POTENTIAL).  
 FT ACT\_SITE 372 372 NUCLEOPHILE (BY SIMILARITY).  
 FT CONFLICT 405 406 EA -> GT (IN REF. 1).  
 FT CONFLICT 428 428 S -> C (IN REF. 1).  
 FT CONFLICT 455 456 RK -> HR (IN REF. 1).  
 SQ SEQUENCE 474 AA; 53935 MW; 02ACE6BEBF211011 CRC64;

Query Match 26.6%; Score 51; DB 1; Length 474;  
 Best Local Similarity 43.3%; Pred. No. 11;  
 Matches 13; Conservative 3; Mismatches 14; Indels 0; Gaps 0;

Qy 1 LGTFWGDTLNCWMLSAFSRYARCLAEGHDG 30  
 | | : | | : | | | | | | | |  
 Db 141 LVTEYGSWRNRKLVEFFSRYARTCFEAFDG 170

# RESULT 7

## EGF\_HUMAN

ID EGF\_HUMAN STANDARD; PRT; 1207 AA.  
 AC P01133;  
 DT 21-JUL-1986 (Rel. 01, Created)  
 DT 13-AUG-1987 (Rel. 05, Last sequence update)  
 DT 15-SEP-2003 (Rel. 42, Last annotation update)  
 DE Pro-epidermal growth factor precursor (EGF) [Contains: Epidermal growth factor (Urogastrone)].  
 GN EGF.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Kidney;  
 RX MEDLINE=87066721; PubMed=3491360;  
 RA Bell G.I., Fong N.M., Stempien M.M., Wormsted M.A., Caput D.,  
 RA Ku L., Urdea M.S., Rall L.B., Sanchez-Pescador R.;  
 RT "Human epidermal growth factor precursor: cDNA sequence, expression  
 RT in vitro and gene organization.";  
 RL Nucleic Acids Res. 14:8427-8446(1986).  
 RN [2]  
 RP SEQUENCE OF 971-1023.  
 RX MEDLINE=77117897; PubMed=300079;  
 RA Gregory H., Preston B.M.;  
 RT "The primary structure of human urogastrone.";  
 RL Int. J. Pept. Protein Res. 9:107-118(1977).  
 RN [3]  
 RP SEQUENCE OF 971-1023.  
 RX MEDLINE=89391964; PubMed=2789514;  
 RA Furuya M., Akashi S., Hirayama K.;  
 RT "The primary structure of human EGF produced by genetic engineering,  
 RT studied by high-performance tandem mass spectrometry.";  
 RL Biochem. Biophys. Res. Commun. 163:1100-1106(1989).  
 RN [4]

RP STRUCTURE BY NMR OF EGF.  
 RX MEDLINE=92395667; PubMed=1522591;  
 RA Hommel U., Harvey T.S., Driscoll P.C., Campbell I.D.;  
 RT "Human epidermal growth factor. High resolution solution structure  
 RT and comparison with human transforming growth factor alpha.";  
 RL J. Mol. Biol. 227:271-282(1992).  
 CC -!- FUNCTION: THE GROWTH FACTOR STIMULATES THE GROWTH OF VARIOUS  
 CC EPIDERMAL AND EPITHELIAL TISSUES IN VIVO AND IN VITRO AND OF SOME  
 CC FIBROBLASTS IN CELL CULTURE.  
 CC -!- SUBCELLULAR LOCATION: Type I membrane protein.  
 CC -!- SIMILARITY: Contains 9 EGF-like domains.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; X04571; CAA28240.1; -.  
 DR PIR; A25531; EGHU.  
 DR PDB; 1IVO; 16-OCT-02.  
 DR PDB; 1JL9; 18-DEC-02.  
 DR Genew; HGNC:3229; EGF.  
 DR MIM; 131530; -.  
 DR GO; GO:0005886; C:plasma membrane; TAS.  
 DR GO; GO:0005155; F:epidermal growth factor receptor activating. . .; TAS.  
 DR GO; GO:0000187; P:activation of MAPK; TAS.  
 DR GO; GO:0007001; P:chromosome organization and biogenesis (sen. . .; TAS.  
 DR GO; GO:0006260; P:DNA replication; TAS.  
 DR InterPro; IPR000152; Asx\_hydroxyl.  
 DR InterPro; IPR001881; EGF\_Ca.  
 DR InterPro; IPR006209; EGF\_like.  
 DR InterPro; IPR000033; Ldl\_receptor\_rep.  
 DR Pfam; PF00008; EGF; 9.  
 DR Pfam; PF00058; ldl\_recept\_b; 7.  
 DR SMART; SM00179; EGF\_CA; 2.  
 DR SMART; SM00135; LY; 8.  
 DR PROSITE; PS00010; ASX\_HYDROXYL; 3.  
 DR PROSITE; PS00022; EGF\_1; 1.  
 DR PROSITE; PS01186; EGF\_2; 7.  
 DR PROSITE; PS01187; EGF\_CA; 3.  
 KW EGF-like domain; Repeat; Growth factor; Transmembrane; Glycoprotein;  
 KW Signal; Polymorphism; 3D-structure.  
 FT SIGNAL 1 22 POTENTIAL.  
 FT CHAIN 23 1207 PRO-EPIDERMAL GROWTH FACTOR.  
 FT CHAIN 971 1023 EPIDERMAL GROWTH FACTOR.  
 FT DOMAIN 23 1032 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 1033 1053 POTENTIAL.  
 FT DOMAIN 1054 1207 CYTOPLASMIC (POTENTIAL).  
 FT DOMAIN 314 355 EGF-LIKE 1.  
 FT DOMAIN 356 396 EGF-LIKE 2, CALCIUM-BINDING (POTENTIAL).  
 FT DOMAIN 397 437 EGF-LIKE 3.  
 FT DOMAIN 435 477 EGF-LIKE 4.  
 FT DOMAIN 741 781 EGF-LIKE 5.  
 FT DOMAIN 831 869 EGF-LIKE 6.

FT	DOMAIN	870	911	EGF-LIKE 7, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	912	952	EGF-LIKE 8, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	972	1013	EGF-LIKE 9.
FT	DISULFID	318	330	BY SIMILARITY.
FT	DISULFID	325	339	BY SIMILARITY.
FT	DISULFID	341	354	BY SIMILARITY.
FT	DISULFID	360	371	BY SIMILARITY.
FT	DISULFID	367	380	BY SIMILARITY.
FT	DISULFID	382	395	BY SIMILARITY.
FT	DISULFID	401	412	BY SIMILARITY.
FT	DISULFID	408	421	BY SIMILARITY.
FT	DISULFID	423	436	BY SIMILARITY.
FT	DISULFID	439	451	BY SIMILARITY.
FT	DISULFID	447	461	BY SIMILARITY.
FT	DISULFID	463	476	BY SIMILARITY.
FT	DISULFID	745	756	BY SIMILARITY.
FT	DISULFID	752	765	BY SIMILARITY.
FT	DISULFID	767	780	BY SIMILARITY.
FT	DISULFID	835	846	BY SIMILARITY.
FT	DISULFID	840	855	BY SIMILARITY.
FT	DISULFID	857	868	BY SIMILARITY.
FT	DISULFID	874	888	BY SIMILARITY.
FT	DISULFID	881	897	BY SIMILARITY.
FT	DISULFID	899	910	BY SIMILARITY.
FT	DISULFID	916	929	BY SIMILARITY.
FT	DISULFID	923	938	BY SIMILARITY.
FT	DISULFID	940	951	BY SIMILARITY.
FT	DISULFID	976	990	
FT	DISULFID	984	1001	
FT	DISULFID	1003	1012	
FT	CARBOHYD	38	38	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	104	104	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	117	117	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	148	148	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	324	324	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	404	404	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	596	596	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	815	815	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	926	926	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	VARIANT	708	708	I -> M.
FT				/FTId=VAR_002275.
SQ	SEQUENCE	1207 AA;	133945 MW;	D627DC828EF782E9 CRC64;

Query Match 26.6%; Score 51; DB 1; Length 1207;  
 Best Local Similarity 56.2%; Pred. No. 29;  
 Matches 9; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

Qy 18 SRYARCLAEGHDGPTQ 33  
 | |||::|| | |  
 Db 841 SMYARCISEGEDATCQ 856

RESULT 8  
 MATK\_EICCR  
 ID MATK\_EICCR STANDARD; PRT; 504 AA.  
 AC Q9GHB1;  
 DT 15-SEP-2003 (Rel. 42, Created)

DT 15-SEP-2003 (Rel. 42, Last sequence update)  
 DT 15-SEP-2003 (Rel. 42, Last annotation update)  
 DE Maturase K (Intron maturase).  
 GN MATK.  
 OS Eichhornia crassipes (Water hyacinth).  
 OG Chloroplast.  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; Liliopsida; Commelinales;  
 OC Pontederiaceae; Eichhornia.  
 OX NCBI\_TaxID=44947;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Fuse S., Tamura M.N.;  
 RT "A phylogenetic analysis of the plastid matK gene with emphasis on  
 RT Melanthiaceae sensu lato.";  
 RL Plant Biol. 2:415-427(2000).  
 CC -!- FUNCTION: Probably assists in splicing chloroplast group II  
 CC introns (By similarity).  
 CC -!- SIMILARITY: BELONGS TO THE INTRON MATURASE FAMILY 2. MATK  
 CC SUBFAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; AB040212; BAB16820.2; -.  
 DR InterPro; IPR000442; Intron\_maturase2.  
 DR Pfam; PF01348; Intron\_maturase2; 1.  
 DR Pfam; PF01824; MatK\_N; 1.  
 KW Chloroplast.  
 SQ SEQUENCE 504 AA; 60109 MW; C68FC55E4AF4C2A7 CRC64;

Query Match 26.0%; Score 50; DB 1; Length 504;  
 Best Local Similarity 31.0%; Pred. No. 17;  
 Matches 9; Conservative 7; Mismatches 13; Indels 0; Gaps 0;

Qy 5 WGDTLNCWMLSAFSRYARCLAEGHDGPTQ 33  
 | | :| ::| | | | : | | ::  
 Db 391 WTDLSDCDIINRFGRICRNLSHYHSGSSK 419

#### RESULT 9

##### GBA2\_NEUCR

ID GBA2\_NEUCR STANDARD; PRT; 355 AA.  
 AC Q05424; Q9URK0;  
 DT 01-OCT-1994 (Rel. 30, Created)  
 DT 01-OCT-1994 (Rel. 30, Last sequence update)  
 DT 15-SEP-2003 (Rel. 42, Last annotation update)  
 DE Guanine nucleotide-binding protein alpha-2 subunit (GP2-alpha).  
 GN GNA-2 OR B11H7.130.  
 OS Neurospora crassa.  
 OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;  
 OC Sordariomycetidae; Sordariales; Sordariaceae; Neurospora.

OX NCBI\_TaxID=5141;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=74-OR23-1A / FGSC 987;  
 RX MEDLINE=93315452; PubMed=8325859;  
 RA Borkovich K.A., Turner G.E.;  
 RT "Identification of a G protein alpha subunit from *Neurospora crassa*  
 RT that is a member of the Gi family.";  
 RL J. Biol. Chem. 268:14805-14811(1993).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=74-OR23-1A / FGSC 987;  
 RX MEDLINE=97432794; PubMed=9286674;  
 RA Baasiri R.A., Lu X., Rowley P.S., Turner G.E., Borkovich K.A.;  
 RT "Overlapping functions for two G protein alpha subunits in *Neurospora*  
 RT *crassa*.";  
 RL Genetics 147:137-145(1997).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=74-OR23-1A / FGSC 987;  
 RX PubMed=12655011;  
 RA Mannhaupt G., Montrone C., Haase D., Mewes H.-W., Aign V.,  
 RA Hoheisel J.D., Fartmann B., Nyakatura G., Kempken F., Maier J.,  
 RA Schulte U.;  
 RT "What's in the genome of a filamentous fungus? Analysis of the  
 RT *Neurospora* genome sequence.";  
 RL Nucleic Acids Res. 31:1944-1954(2003).  
 CC -!- FUNCTION: GUANINE NUCLEOTIDE-BINDING PROTEINS (G PROTEINS) ARE  
 CC INVOLVED AS MODULATORS OR TRANSDUCERS IN VARIOUS TRANSMEMBRANE  
 CC SIGNALING SYSTEMS.  
 CC -!- SUBUNIT: G proteins are composed of 3 units (alpha, beta and  
 CC gamma).  
 CC THE ALPHA CHAIN CONTAINS THE GUANINE NUCLEOTIDE BINDING SITE.  
 CC -!- SIMILARITY: BELONGS TO THE G-ALPHA FAMILY. SUBFAMILY 3 (G(Q)).  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; L11452; AAA02559.1; -.  
 DR EMBL; AF004846; AAD01207.1; -.  
 DR EMBL; BX294092; CAD71253.1; -.  
 DR PIR; T50479; T50479.  
 DR InterPro; IPR001019; Gprotein\_alpha.  
 DR Pfam; PF00503; G-alpha; 1.  
 DR PRINTS; PR00318; GPROTEINA.  
 DR ProDom; PD000281; Gprotein\_alpha; 1.  
 DR SMART; SM00275; G\_alpha; 1.  
 KW GTP-binding; Transducer; Multigene family.  
 FT NP\_BIND 41 48 GTP (BY SIMILARITY).  
 FT NP\_BIND 201 205 GTP (BY SIMILARITY).  
 FT NP\_BIND 270 273 GTP (BY SIMILARITY).  
 FT CONFLICT 19 20 EL -> DV (IN REF. 2).

SQ SEQUENCE 355 AA; 41361 MW; 61733B89EABB7409 CRC64;

Query Match 25.8%; Score 49.5; DB 1; Length 355;  
Best Local Similarity 44.0%; Pred. No. 14;  
Matches 11; Conservative 4; Mismatches 9; Indels 1; Gaps 1;

Qy 7 DTLNCWM-LSAFSRYARCLAEHGDG 30  
: :|| : | | | :|| | ||  
Db 217 ENVNCLLFLVAISGYDQCLVEDKDG 241

RESULT 10

ENV\_HV1S1

ID ENV\_HV1S1 STANDARD; PRT; 847 AA.

AC P19550;

DT 01-FEB-1991 (Rel. 17, Created)

DT 01-FEB-1991 (Rel. 17, Last sequence update)

DT 15-SEP-2003 (Rel. 42, Last annotation update)

DE Envelope polyprotein GP160 precursor [Contains: Exterior membrane  
glycoprotein (GP120); Transmembrane glycoprotein (GP41)].

GN ENV.

OS Human immunodeficiency virus type 1 (SF162 isolate) (HIV-1).

OC Viruses; Retrovird viruses; Retroviridae; Lentivirus.

OX NCBI\_TaxID=11691;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=90347835; PubMed=2384920;

RA Cheng-Mayer C., Quiroga M., Tung J.W., Dina D., Levy J.;

RT "Viral determinants of human immunodeficiency virus type 1 T-cell or  
macrophage tropism, cytopathogenicity, and CD4 antigen modulation.";

RL J. Virol. 64:4390-4398(1990).

CC

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC

DR EMBL; M65024; AAA45072.1; -.

DR PDB; 1OBE; 15-MAY-97.

DR HIV; M38428; ENV\$SF162.

DR InterPro; IPR000328; Env\_GP41.

DR InterPro; IPR000777; GP120.

DR Pfam; PF00516; GP120; 1.

DR Pfam; PF00517; GP41; 1.

KW AIDS; Coat protein; Polyprotein; Glycoprotein; Transmembrane; Signal;  
KW 3D-structure.

FT SIGNAL 1 29

FT CHAIN 30 502 EXTERIOR MEMBRANE GLYCOPROTEIN.

FT CHAIN 503 847 TRANSMEMBRANE GLYCOPROTEIN.

FT DISULFID 53 73 BY SIMILARITY.

FT DISULFID 118 203 BY SIMILARITY.

FT DISULFID 125 194 BY SIMILARITY.

FT DISULFID 130 155 BY SIMILARITY.

FT DISULFID 216 245 BY SIMILARITY.

FT	DISULFID	226	237	BY SIMILARITY.
FT	DISULFID	294	328	BY SIMILARITY.
FT	DISULFID	374	435	BY SIMILARITY.
FT	DISULFID	381	408	BY SIMILARITY.
FT	CARBOHYD	87	87	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	135	135	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	154	154	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	186	186	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	195	195	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	232	232	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	239	239	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	260	260	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	274	274	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	293	293	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	299	299	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	329	329	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	336	336	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	352	352	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	382	382	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	388	388	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	392	392	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	398	398	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	401	401	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	438	438	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	454	454	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	602	602	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	607	607	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	616	616	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	628	628	N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ	SEQUENCE	847 AA; 96135 MW; 0A901317FD7FF2AB CRC64;		

Query Match 25.8%; Score 49.5; DB 1; Length 847;  
 Best Local Similarity 33.3%; Pred. No. 33;  
 Matches 11; Conservative 5; Mismatches 10; Indels 7; Gaps 1;

Qy 4 FWGDTLNCWM-----LSAFSRYARCLAEAGHD 29  
 :||: | |: :| | | :||| |  
 Db 786 YWGNLLQYWIQELKNSAVSLFDAIAIAVAEGTD 818

RESULT 11  
 APAH\_BURMA  
 ID APAH\_BURMA STANDARD; PRT; 282 AA.  
 AC Q9AEV8;  
 DT 28-FEB-2003 (Rel. 41, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Bis(5'-nucleosyl)-tetraphosphatase, symmetrical (EC 3.6.1.41)  
 DE (Diadenosine tetraphosphatase) (Ap4A hydrolase) (Diadenosine 5',5''-  
 DE P1,P4-tetraphosphate pyrophosphohydrolase).  
 GN APAH.  
 OS Burkholderia mallei (Pseudomonas mallei).  
 OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;  
 OC Burkholderiaceae; Burkholderia.  
 OX NCBI\_TaxID=13373;  
 RN [1]  
 RP SEQUENCE FROM N.A.



RA Burtnick M.N., Brett P.J., Woods D.E.;  
 RT "Physical and molecular characterization of lipopolysaccharide  
 RT O-antigens from Burkholderia mallei."  
 RL Submitted (MAR-2001) to the EMBL/GenBank/DDBJ databases.  
 CC -!- FUNCTION: Hydrolyzes diadenosine 5',5'''-P1,P4-tetraphosphate to  
 CC yield ADP (By similarity).  
 CC -!- CATALYTIC ACTIVITY: P(1),P(4)-bis(5'-adenosyl)tetraphosphate +  
 CC H(2)O = 2 ADP.  
 CC -!- SIMILARITY: Belongs to the Ap4A hydrolase family.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; AY028370; AAK27390.1; -.  
 DR HAMAP; MF\_00199; -; 1.  
 DR InterPro; IPR004617; ApaH.  
 DR InterPro; IPR004843; M-ppestrase.  
 DR InterPro; IPR006186; T\_phtase\_apah.  
 DR Pfam; PF00149; Metallophos; 1.  
 DR ProDom; PD000252; T\_phtase\_apah; 1.  
 DR TIGRFAMs; TIGR00668; apah; 1.  
 KW Hydrolase.  
 SQ SEQUENCE 282 AA; 30631 MW; 7F83BE3404103374 CRC64;

Query Match 25.5%; Score 49; DB 1; Length 282;  
 Best Local Similarity 30.2%; Pred. No. 13;  
 Matches 13; Conservative 2; Mismatches 8; Indels 20; Gaps 2;

Qy 5 WGDTL-----NCW-----MLSAFSRYARCLAEG 27  
 | ||| ||| :||:| | ||  
 Db 151 WRDTLRSLYGNPNCWSPDLKHADRLRVAFNAFTRIRFCTPEG 193

# RESULT 12

## APAH\_BURPS

ID APAH\_BURPS STANDARD; PRT; 282 AA.  
 AC O69115;  
 DT 30-MAY-2000 (Rel. 39, Created)  
 DT 30-MAY-2000 (Rel. 39, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Bis(5'-nucleosyl)-tetraphosphatase, symmetrical (EC 3.6.1.41)  
 DE (Diadenosine tetraphosphatase) (Ap4A hydrolase) (Diadenosine 5',5'''-  
 DE P1,P4-tetraphosphate pyrophosphohydrolase).  
 GN APAH.  
 OS Burkholderia pseudomallei (Pseudomonas pseudomallei).  
 OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;  
 OC Burkholderiaceae; Burkholderia.  
 OX NCBI\_TaxID=28450;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=1026b;  
 RA DeShazer D., Brett P.J., Woods D.E.;

RT "The type II O-antigen moiety of Burkholderia pseudomallei  
RT lipopolysaccharide is required for serum resistance and virulence.";  
RL Submitted (MAY-1998) to the EMBL/GenBank/DDBJ databases.  
CC -!- FUNCTION: Hydrolyzes diadenosine 5',5'''-P1,P4-tetraphosphate to  
CC yield ADP (By similarity).  
CC -!- CATALYTIC ACTIVITY: P(1),P(4)-bis(5'-adenosyl)tetraphosphate +  
CC H(2)O = 2 ADP.  
CC -!- SIMILARITY: Belongs to the Ap4A hydrolase family.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; AF064070; AAD05453.1; -.  
DR HAMAP; MF\_00199; -; 1.  
DR InterPro; IPR004617; ApaH.  
DR InterPro; IPR004843; M-ppetrase.  
DR InterPro; IPR006186; T\_phtase\_apah.  
DR Pfam; PF00149; Metallophos; 1.  
DR ProDom; PD000252; T\_phtase\_apah; 1.  
DR TIGRFAMs; TIGR00668; apah; 1.  
KW Hydrolase.  
SQ SEQUENCE 282 AA; 30609 MW; 5D8BF833C5C27F44 CRC64;

Query Match 25.5%; Score 49; DB 1; Length 282;  
Best Local Similarity 30.2%; Pred. No. 13;  
Matches 13; Conservative 2; Mismatches 8; Indels 20; Gaps 2;

Qy 5 WGDTL-----NCW-----MLSAFSRYARCLAEG 27  
| | | | | : | | | | |  
Db 151 WRDTLRSLYGNDPNCWSPDLKHADRLRVAFNAFTRIRFCTPEG 193

# RESULT 13

## PEPE\_ECOL6

ID PEPE\_ECOL6 STANDARD; PRT; 229 AA.  
AC Q8FB55;  
DT 15-SEP-2003 (Rel. 42, Created)  
DT 15-SEP-2003 (Rel. 42, Last sequence update)  
DT 15-SEP-2003 (Rel. 42, Last annotation update)  
DE Peptidase E (EC 3.4.13.21) (Alpha-aspartyl dipeptidase) (Asp-specific  
DE dipeptidase) (Dipeptidase E).  
GN PEPE OR C4980.  
OS Escherichia coli O6.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
OC Enterobacteriaceae; Escherichia.  
OX NCBI\_TaxID=217992;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=O6:H1 / CFT073 / ATCC 700928;  
RX MEDLINE=22388234; PubMed=12471157;  
RA Welch R.A., Burland V., Plunkett G. III, Redford P., Roesch P.,  
RA Rasko D., Buckles E.L., Liou S.-R., Boutin A., Hackett J., Stroud D.,

RA Mayhew G.F., Rose D.J., Zhou S., Schwartz D.C., Perna N.T.,  
 RA Mobley H.L.T., Donnenberg M.S., Blattner F.R.;  
 RT "Extensive mosaic structure revealed by the complete genome sequence  
 RT of uropathogenic *Escherichia coli*.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:17020-17024(2002).  
 CC -!- FUNCTION: Hydrolyzes dipeptides containing N-terminal aspartate  
 CC residues. May play a role in allowing the cell to use peptide  
 CC aspartate to spare carbon otherwise required for the synthesis of  
 CC the aspartate family of amino acids (By similarity).  
 CC -!- CATALYTIC ACTIVITY: Dipeptidase E catalyzes the hydrolysis of  
 CC dipeptides Asp-Xaa. It does not act on peptides with N-terminal  
 CC Glu, Asn or Gln, nor does it cleave isoaspartyl peptides.  
 CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).  
 CC -!- SIMILARITY: Belongs to peptidase family S51.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; AE016770; AAN83406.1; -.  
 DR HAMAP; MF\_00510; -; 1.  
 DR Pfam; PF03575; Peptidase\_S51; 1.  
 KW Hydrolase; Serine protease; Dipeptidase; Complete proteome.  
 FT ACT\_SITE 120 120 CHARGE RELAY SYSTEM (BY SIMILARITY).  
 FT ACT\_SITE 135 135 CHARGE RELAY SYSTEM (BY SIMILARITY).  
 FT ACT\_SITE 157 157 CHARGE RELAY SYSTEM (BY SIMILARITY).  
 SQ SEQUENCE 229 AA; 24560 MW; 519FB4356C843CC2 CRC64;

Query Match 25.0%; Score 48; DB 1; Length 229;  
 Best Local Similarity 37.0%; Pred. No. 14;  
 Matches 10; Conservative 3; Mismatches 14; Indels 0; Gaps 0;

Qy 7 DTLNCWMLSAFSRYARCLAEGHDGPTQ 33  
 | | | : | : | | | | | :  
 Db 145 DALNLFPLQINPHFTNALPEGHKGETR 171

RESULT 14  
 PEPE\_ECOLI  
 ID PEPE\_ECOLI STANDARD; PRT; 229 AA.  
 AC P32666;  
 DT 01-OCT-1993 (Rel. 27, Created)  
 DT 01-OCT-1993 (Rel. 27, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Peptidase E (EC 3.4.13.21) (Alpha-aspartyl dipeptidase) (Asp-specific  
 DE dipeptidase) (Dipeptidase E).  
 GN PEPE OR B4021 OR Z5612 OR ECS4939.  
 OS *Escherichia coli*, and  
 OS *Escherichia coli* O157:H7.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
 OC Enterobacteriaceae; *Escherichia*.  
 OX NCBI\_TaxID=562, 83334;  
 RN [1]

RP SEQUENCE FROM N.A.  
 RC STRAIN=K12 / MG1655;  
 RX MEDLINE=94089392; PubMed=8265357;  
 RA Blattner F.R., Burland V.D., Plunkett G. III, Sofia H.J.,  
 RA Daniels D.L.;  
 RT "Analysis of the Escherichia coli genome. IV. DNA sequence of the  
 RT region from 89.2 to 92.8 minutes.";  
 RL Nucleic Acids Res. 21:5408-5417(1993).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=O157:H7 / EDL933 / ATCC 700927;  
 RX MEDLINE=21074935; PubMed=11206551;  
 RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,  
 RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,  
 RA Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,  
 RA Grotbeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamousis K.,  
 RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,  
 RA Welch R.A., Blattner F.R.;  
 RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7.";  
 RL Nature 409:529-533(2001).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=O157:H7 / RIMD 0509952;  
 RX MEDLINE=21156231; PubMed=11258796;  
 RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,  
 RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,  
 RA Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,  
 RA Kuhara S., Shiba T., Hattori M., Shinagawa H.;  
 RT "Complete genome sequence of enterohemorrhagic Escherichia coli  
 RT O157:H7 and genomic comparison with a laboratory strain K-12.";  
 RL DNA Res. 8:11-22(2001).  
 CC -!- FUNCTION: Hydrolyzes dipeptides containing N-terminal aspartate  
 CC residues. May play a role in allowing the cell to use peptide  
 CC aspartate to spare carbon otherwise required for the synthesis of  
 CC the aspartate family of amino acids.  
 CC -!- CATALYTIC ACTIVITY: Dipeptidase E catalyzes the hydrolysis of  
 CC dipeptides Asp|-Xaa. It does not act on peptides with N-terminal  
 CC Glu, Asn or Gln, nor does it cleave isoaspartyl peptides.  
 CC -!- SUBCELLULAR LOCATION: Cytoplasmic.  
 CC -!- SIMILARITY: Belongs to peptidase family S51.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; U00006; AAC43115.1; -.  
 DR EMBL; AE000475; AAC76991.1; -.  
 DR EMBL; AE005634; AAG59213.1; -.  
 DR EMBL; AP002567; BAB38362.1; -.  
 DR PIR; A86094; A86094.  
 DR PIR; C91246; C91246.  
 DR PIR; D65209; D65209.  
 DR HSSP; P36936; 1FYE.

DR MEROPS; S51.001; -.  
 DR EcoGene; EG11920; pepE.  
 DR HAMAP; MF\_00510; -; 1.  
 DR InterPro; IPR005320; Peptidase\_S51.  
 DR Pfam; PF03575; Peptidase\_S51; 1.  
 KW Hydrolase; Serine protease; Dipeptidase; Complete proteome.  
 FT ACT\_SITE 120 120 CHARGE RELAY SYSTEM (BY SIMILARITY).  
 FT ACT\_SITE 135 135 CHARGE RELAY SYSTEM (BY SIMILARITY).  
 FT ACT\_SITE 157 157 CHARGE RELAY SYSTEM (BY SIMILARITY).  
 SQ SEQUENCE 229 AA; 24570 MW; 53D4D8395DFC63FD CRC64;

Query Match 25.0%; Score 48; DB 1; Length 229;  
 Best Local Similarity 37.0%; Pred. No. 14;  
 Matches 10; Conservative 3; Mismatches 14; Indels 0; Gaps 0;

Qy 7 DTLNCWMLSAFSRYARCLAEHGDGPTQ 33  
 | | | : | : | | | | | :  
 Db 145 DALNLFPLQINPHFTNALPEGHKGETR 171

# RESULT 15

## HEAD\_BPGA1

ID HEAD\_BPGA1 STANDARD; PRT; 472 AA.  
 AC Q9FZW7;  
 DT 28-FEB-2003 (Rel. 41, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Major head protein.  
 GN 8.  
 OS Bacteriophage GA-1.  
 OC Viruses; dsDNA viruses, no RNA stage; Caudovirales; Podoviridae;  
 OC phi-29-like viruses.  
 OX NCBI\_TaxID=12345;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Meijer W.J.J., Horcajadas J.A., Salas M.;  
 RT "The phi29 family of phages."  
 RL Submitted (DEC-2000) to the EMBL/GenBank/DDBJ databases.

CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----

DR EMBL; X96987; CAC21529.1; -.  
 SQ SEQUENCE 472 AA; 53022 MW; 3104821153B1C4C2 CRC64;  
 Query Match 25.0%; Score 48; DB 1; Length 472;  
 Best Local Similarity 33.3%; Pred. No. 30;  
 Matches 8; Conservative 6; Mismatches 10; Indels 0; Gaps 0;

Qy 2 GTFWGDTLNCWMLSAFSRYARCLA 25  
 | : | | : | : : || : | : |  
 Db 318 GMYWNYLHVWQVLSTSRFANAVA 341  
 Search completed: January 30, 2004, 11:25:07  
 Job time : 2.28405 secs